RADC-TDR-63-324

PROCEEDINGS OF SYMPOSIUM ON HUMAN FACTOR ASPECTS OF PHOTO INTERPRETATION

TECHNICAL DOCUMENTARY REPORT NO. RADC-TDR-63-324

September 1963

Human Engineering Laboratory
Rome Air Development Center
Research and Technology Division
Air Force Systems Command
Griffiss Air Force Base, New York

Project 6244

ABSTRACT

This report is an edited record of a symposium entitled "Human Factor Aspects of Photo Interpretation" held on 7-8 November 1962, sponsored by Rome Air Development Center (RADC).

Five RADC contractors working simultaneously on specific problems in the area were invited to participate in conjunction with representatives of other interested military and industrial organizations. The prime goal of the symposium was the interchange of ideas and information relative to ongoing research efforts to improve the efficiency of the military interpreter. Presentations and discussions in the areas of image quality factors, comparative cover analysis, temporal factors and interpreter training are reported.

Dr. Shelton MacLeod of the Human Engineering Laboratory, RADC, was chairman of the symposium.

PUBLICATION REVIEW

This report has been reviewed and is approved.

Approved: Carlo Crocetti

CARLO P. CROCETTI
Chief, Human Engineering Laboratory
Directorate of Engineering

Approved: Joseph Mannaby

WILLIAM P. BETHKE
Director of Engineering

FOR THE COMMANDER: Irving J. Gabelman

IRVING J. GABELMAN
Director of Advanced Studies

AIR FORCE, Cato Show Prtg. Co. 10/18/63 - 181 -

TABLE OF CONTENTS

AD-422 432

Contents *Page*

I. INTRODUCTORY DISCUSSIONS

Opening Address 1

The Automation of Photo Interpreter Functions 3

Human Factor Problems: Objectives and Approaches 5

II. IMAGE QUALITY FACTORS

Measurement and Evaluation of Ground Resolution 9

Quality Categorization of Aerial Photography 26

Psychophysical Aspects of Image Quality 42

A Study of Image Qualities and Speeded Intrinsic Target Recognition 53

III. COMPARATIVE COVER ANALYSIS

Factors Affecting Change Discrimination 61

IV. TEMPORAL FACTORS

Research on the Relationship Between Time and PI Performance 87

Application of Rapid Presentation Techniques in Photo Interpretation 109

V. INTERPRETER TRAINING

Photo Interpretation Courses Taught at Sheppard AFB, Texas 123

VI. CONCLUDING DISCUSSIONS

Other Human Factors Problem Areas Requiring Study 127

Concluding Remarks 136

ATTENDANCE LIST 137

SECTION I

INTRODUCTORY DISCUSSIONS

OPENING ADDRESS

Dr. Carlo P. Crocetti, Chief
Dr. Shelton MacLeod, Research Psychologist
Human Engineering Laboratory
Rome Air Development Center

Dr. Crocetti:

Rome Air Development Center (RADC) has had a long-time interest in the R&D aspects of photo interpretation. This interest ranges from an attempt toward a basic understanding of the fundamental processes which underlie interpretation functions to even more mundane kinds of evaluations of techniques for aiding an interpreter in doing a very specific function. This has led to a broad program at RADC; what underlies this particular symposium is the fact that we, as a single center, have a large number of contractors all working simultaneously on various problems in this area. We felt a number of months ago that it would be desirable to bring these people together to discuss common problems. There are problems of experimental control and data collection that are common to all of us who have an interest in photo interpretation and the interchange of such data should be of mutual benefit. As we considered this objective, we recognized that other people, in addition to RADC and our own contractors, are also working in this area; and, rather than wait for publication of reports (which in many cases are as long as a year or two away), they might also wish to participate in such a meeting. I think the large number of organizations that are represented this morning is an excellent indication of the interest in this field. We have deliberately held attendance by invitation only because it was our desire to keep the number of attendees to a reasonable level. We felt that if we had over two hundred participants we would no longer have the aura of a symposium. In fact, we are probably pushing our limits now, and whether we really have the right to call this a symposium, we hope that we can be successful in this interchange. I think you will notice that after every presentation, there will be an opportunity for group discussion. I would like to encourage each of you to participate because we feel that this is the prime objective of this meeting. This is not a technical session of a paper-presenting type. It is actually, we hope, an interchange of ideas for all of us. I would like to again thank each of you for coming because we feel that you represent an interest and source of information to those of us interested in this area.

Dr. MacLeod:

We are here to exchange and coordinate information of a scientific and technical nature which is concerned with improving the efficiency of the military interpreter. I think, perhaps, the best thing which can result from this symposium is that, by our presentations and discussions, we can raise the level of sophistication in thinking of more effective ways to plan and conduct badly needed human factors research in this key area. The formal presentations will be given by five contractors and, in addition, we are fortunate enough to have representatives from the Army, Air Training Command, and International Business Machine Company, who will provide additional presentations. At this time I would like to introduce Mr. Lynwood Sinnamon of the Information Processing Laboratory, who will give you some background on RADC programs for automating photo interpretation research.

THE AUTOMATION OF PHOTO INTERPRETER FUNCTIONS

Lynwood D. Sinnamon, Jr.
Information Processing Laboratory
Rome Air Development Center
Griffiss Air Force Base, N.Y.

ABSTRACT

Background information is presented regarding the contribution of automation to photo interpretation. Projects in optical processing, enhancement, textural analysis, etc. are discussed. Some indications of probable trends and needs for future research are noted.

The Reconnaissance and Interpretation Branch of the RADC Information Processing Laboratory is not only interested in the study of the human interpreter as a system element, as demonstrated by our financial support of five contractors here today and our moral support of many more, but also in the development of techniques that will assist the analyst in his pattern recognition tasks and, ultimately, in his analytical tasks. This, of course, will not be an overnight occurrence and, because of the extreme complexity of some of the tasks, may never be completely automated.

There has been an abundance of articles on artificial intelligence, "bionics," and pattern recognition. However, there have been very few developments that could be applied to Air Force problems. The principal fault with most of these is that they were developed for high contrast characters with simple shapes. I will not attempt to discuss all of the techniques because there have been many summary articles covering the field, but I will cover those techniques that have been supported by RADC because we felt they might have specific application to Air Force problems.

Our earliest support was for optical processing because it was most easily instrumented and studied. Optical properties such as cross-correlation with specific and generalized shapes, two-channel area matching, and aperture filtering, were studied. In the output from such processing, one looks for the appearance of a strong indicator for the presence or absence of the desired target. However, during the early optical processing work, no such indicator was found. Independent research has revealed that using an optical property, such as the power spectrum, as a filter made it possible to detect objects by "enhancing" them; but a compromise was always involved when the structure of the original had to be maintained. RADC has been supporting research to apply such phenomena to locating targets even if the optical image becomes distorted. The results were very encouraging because many targets were found that could be quickly highlighted. It should be remembered that these findings were the result of well-controlled laboratory experiments with high-quality photographs. It is now necessary to analyze these results more thoroughly as functions of various image variables. A study is to be conducted to determine if the highlighting of targets would be useful for an interpreter.

Optical processing has been preferred because it retains the imagery in its original form. But there are important uses for electronic processing of which the Semiautomatic Target Recognition Device, the Automatic Target Recognition Device and the Textural Analysis Work of Budd Electronics and Ford Instruments are good examples. These techniques are not designed to recognize a single indicator, as is usually the case in optical processing, but rather to measure several indicators and to use this data in a slightly more complex decision space than the simple binary one used in optical processing. The indicators can range all the way from simple absolute and relative density measurements to something as complex as the average distance between edges summed in all directions.

When we are no longer interested in simple binary decisions based on the presence or absence of a specific indicator, more complicated linear and nonlinear mathematical analysis is required. The widely publicized adaptive memories are simply linear discriminators. To reduce the size of such memories, the data must usually be normalized. There are currently two efforts being supported by RADC to study the techniques for normalizing photographs.

As more indicators are measured, linear discrimination fails and decisions based on nonlinear techniques must be made. It is then possible to measure eight to twelve indicators, and determine the presence or absence of several target types. The nonlinear techniques, developed for use in statistical problems, have been extended during the past couple of years for use in pattern recognition. As mathematically powerful as these approaches appear to be, they by no means are to be considered as replicas of the human decision process.

What is envisioned by us for the future? Up until now, we have studied the human as a system element, observing outputs for a particular input, but never really trying to analyze the operations of the various parts of the human perception apparatus. Such research must begin soon. Almost complete automation of the pattern recognition tasks is also envisioned, although still far off. It is possible to assist the interpreter and the analyst with currently available techniques. Photographs can be screened for clouds, water, desert, unexposed film and straight edges. On a roll of film such data may be useful in directing the interpreter to important frames or probable targets as would be required in first-phase reporting. As more accurate equipment is built, more complicated tasks will be similarly aided. Likewise, quick quality measurements can be made to determine contrast, acutance and other indices which affect the interpretability of a frame.

In conclusion, we are left with a very important but unanswered question: Will all of this automated assistance actually help the interpreter and the analyst? It is hoped that a program can be started to evaluate the effects of automatic assistance to an interpreter. In such a program various tasks would be performed by interpreters with and without simulated automatic assistance. The results should prove to be very instructive.

I believe that automation will become a part of interpretation just as it has become part of almost everything else in contemporary life.

HUMAN FACTOR PROBLEMS: OBJECTIVES AND APPROACHES

Dr. Shelton MacLeod, Research Psychologist
Human Engineering Laboratory
Rome Air Development Center
Griffiss Air Force Base, N.Y.

ABSTRACT

Several key human factor problems in photo interpretation are depicted along with required objectives and approaches. The need for an optimal integration of man and machine phases of the interpretation process is discussed with reference to four major classes of research toward more efficient interpreter performance. Specific criteria of realism and applicability (common requirements of the five contractor efforts) are discussed.

I would like to talk in general about what constitutes the human factor problem in photo interpretation and some of the objectives and approaches that the Human Engineering Laboratory, working in conjunction with the information Processing Laboratory, have considered with this type of problem. It is now obvious how important the photo interpretation function is in the light of the presently existing Cuban crisis, wherein humans looking at photography are required to give us a type of vitally needed intelligence in a timely fashion. I think a very important point to be made is that the role the human plays in photo interpretation is not only very essential but also, in my opinion, one of the less likely functions soon to be taken over by automatic systems. The human provides something unique here which will be very difficult to program in any reliable way into a computer system. He must go through perceptual decision-making routines (looking, searching, associating information, drawing conclusions) which require the kinds of capacities for vision and decision making which humans uniquely possess. The functions he performs involve screening, search, locating targets, identifying targets (the target itself or actually getting at the precise features that make the differences between targets in the same general class). He identifies changes of military significance (often very subtle changes which require a very precise type of reasoning and discrimination).

At the same time that the human is so essential, he probably will have to operate under conditions which will greatly constrain his efficiency. With increased extent and frequency of reconnaissance surveillance, increased volumes of photography will confront him. There will be at the same time an increasing variety of imagery; not only conventional photographic sensors will be flown, but also infrared, side-looking radar, and perhaps even other sensor packages. Timeliness is also being given a greater and greater amount of emphasis. An interpreter will be required to make flash reports within minutes to identify high priority type information or to make reports from collection vehicles where perhaps he has had only one look at the information.

Now, in view of the constraints which are being put upon a man in this case, there is all the more urgency for the human factors research programs to insure that we can maintain adequate efficiency on the part of the interpreter. These programs can follow a number of different approaches. We will mention four classes of programs which can be followed to arrive at more efficient interpretation.

First, we can try to improve the working conditions of the photo interpreter, make better decisions of how we assign tasks to him; arrange for more efficient duty cycles, devise new and improved modes of communication among the interpreters (perhaps between the photo interpreter and automatic information storage systems); improve the design of the equipment of the interpreter and his work space.

Secondly, aside from this human engineering type of consideration, we have a specific concern in improving the way the interpreter utilizes photography and reference material. The question of image enhancement arises here, improving his ability to extract information from degraded photography. Image degradation, of course, is another constraint which I could have mentioned, implied by the fact that modern collection systems will provide us with small scale degraded imagery which is considerably less than the kind of quality which interpreters have been using in the past.

A third type of program involves developing improved equipment aids: better viewing and mensurational devices; the kinds of automatic devices Mr. Sinnamon mentioned, which will enable the interpreter to screen targets more effectively, pinpointing objects which he might have missed, improving his ability to detect targets and changes. Devices also require development which can meter the quality of the photography for either the interpreter or personnel who are handling the photography. The point here, which we will get into later in the symposium, is to be able to specify quality in such a way that it ties in directly with the ability of the interpreter to extract information.

Finally, a very important area has to do with improved methods of selecting and training interpreters, developing reliable test batteries to allow us to pick the most capable people in different specialties of interpretation, along with more adequate training procedures and aids to training.

In support of such objectives RADC has initiated five contract efforts which will be discussed in greater detail. One effort is concerned with measurement and evaluation of ground resolution as a determiner of how well photography lends itself to information extraction. This is being handled by Minneapolis-Honeywell. Another related effort, conducted by Cornell Aeronautical Laboratory, concerns developing and evaluating a summary measure of image quality so that we can appraise photography in terms of its information extraction potential, not only in terms of one image quality variable, but in terms of the joint effects and trade-offs among more than one variable. Another effort, by Boeing, deals with delineation of factors affecting change discrimination. We have a program on temporal factors with Applied Psychology Corporation which will lead to an assessment of the effect of time-limiting on the quality of information extraction. This study will also be concerned with the establishment of more efficient work-rest intervals for the photo interpreter. Finally, we have a study by RCA in the temporal factors area wherein we are going to develop and evaluate a rapid presentation technique for training the interpreter.

In all of these studies we have set up a number of criteria which we feel will make the data more directly applicable.

First of all, we are using as test materials real aerial photography obtained from various sources; or, if we can demonstrate its effectiveness, we can use simulated photography. We have tried to include in the photography representative strategic and tactical targets of general interest throughout the Armed Forces.

We have also attempted to obtain ground truth data (either through outside information of what is actually in the photography, or through the consensus of experienced interpreters viewing the photography under ideal conditions) to determine what is actually in the photography so we can appraise the performance of the subjects.

Subjects who have had military photo interpretation experience will be used. The subjects' responses are to be representative of actual photo interpretation tasks. Also, we have tried to establish enough of the working conditions of the photo interpreter in our tests to achieve adequate realism from that point of view.

The data of the tests in all cases are representative measures of photo interpretation performance; i.e., objective measures of the speed, accuracy or completeness with which the subjects can extract information from the test materials under experimental conditions.

Finally, we have attempted in all of our experiments to study the effects of a principal independent variable (whether this be a time limit, the mode of presenting comparative-cover pairs, or the effect of ground resolution) as it interacts with other significant variables. For example, if we are trying to assess the effect of limiting the viewing time for the interpreter, we will study this as a function of several levels of image quality, the level of experience of the interpreter, the density and detail of the targets present in the photography, etc., so we can obtain results which will be generalizable to a wide variety of operational conditions.

SECTION II

IMAGE QUALITY FACTORS

MEASUREMENT AND EVALUATION OF GROUND RESOLUTION

Robert N. Cook
Fred B. Meeker
Minneapolis - Honeywell Regulator Co.
Duarte, California

ABSTRACT

A study is described which will undertake the analysis of photo interpretation performance as a function of photographic ground resolution and contrast. Interpreter performance is to be measured in terms of accuracy, speed and features used in identification. Twenty-nine types of targets are to be investigated in both real and simulated aerial photography. Results of the study should facilitate prediction of minimum levels of PI performance for photos having a given image quality. Concerning reconnaissance mission planning, the results will show what ground resolution and contrast (neglecting grain effects) are necessary to provide X level of accuracy and Y level of speed in target identification.

MILITARY SIGNIFICANCE OF CONTRACT

The ground resolution study is concerned with gathering enough data to allow the Air Force to make intelligent predictions about photo interpreter performance as a function of image quality. Two traditional image quality factors are being studied - ground resolution and contrast. Grain is not being tested, but an attempt is being made to avoid excessive grain in the experimental imagery.

The military significance of this study is twofold. The data from the study will make it possible to predict minimum levels of photo interpreter performance for photographs having a given image quality. Specifically, for a given target type, ground resolution, and contrast, this study will show what minimum accuracy and speed can be expected in the identification of detected targets by experienced photo interpreters (PI's). This capability will aid photo interpreter work planning. The study data will also aid in photo reconnaissance mission planning. In this case it will be possible to predict minimum levels of image quality required for efficient photo interpretation. Specifically, for a given minimum level of photo interpreter experience and target-type to be identified, this study will show what ground resolution and contrast (neglecting grain effects) are necessary to give X level of accuracy and Y level of speed in target identification.

CONTRACT GOALS

Cost and time limitations make this study a sampling type. The results will apply to less than 40 target types, but these target types have been chosen to be representative of a large population of tactical targets in which the Air Force is interested. Some interpolation between target types tested and similar targets not tested will be possible. With the preceding qualification in mind, the contract goals may be listed as follows:

1. Determine which targets, on selected list of tactical targets, can be identified at various levels of ground resolution and contrast by experienced PI's.

2. Determine the accuracy and speed of PI identification performance at various levels of ground resolution and contrast.

3. Determine which features, from a selected list of geometric features, are preferred by PI's tested in aiding target identification at various levels of ground resolution and contrast.

4. Determine the correlation in accuracy and speed of PI identification between aerial and simulated aerial photographs for each target showing on both types of photography.

The study has both engineering and human factor study areas. Engineering was required to develop techniques to measure photographic ground resolution and contrast. Human factor work was performed to design a series of experiments to evaluate photo interpreter performance. In addition, human factor concepts were used to refine techniques used in measuring ground resolution and contrast.

TECHNICAL APPROACH

The technical approach consists of performing the study in three related, semiconsecutive phases. Some parallel effort between phases is necessary to feedback phase-interdependent ideas. Phase I involved the development of engineering techniques to grade aerial photography. Phase II deals with the procurement and grading of aerial and simulated aerial photographs. The techniques developed in Phase I apply to the grading in Phase II. Phase III is concerned with the design and execution of a series of three photo interpretation experiments. Each of the three phases will be reviewed briefly to show the scope of work.

PHASE I

Most work in Phase I went into the development of a means of measuring film resolution, which with scale, determines ground resolution. Scale measurement was routine in most cases, so this focused attention on the measurement of film resolution. Five techniques for measuring film resolution were hypothesized and investigated. They are:

1. Sharp edge rise time (electronic).
2. Spectral analysis (electronic).
3. Variable scanning spot size (electronic).
4. Useful magnification (optical).
5. Transition bandwidth (optical).

The first three methods listed above employed a flying-spot scanner cathode ray tube system, which transformed the photo transparencies into a video signal suitable for measurement purposes. Opaque prints, therefore, were not used.

Method 1 assumes that a usable relationship exists between sharp edge gradient width, corresponding rise time of the video signal, and the photographic resolution. Actually, the edge gradient width and shape determine the film resolution over most of the normal contrast range (contrast ratios of 2/1 up to 30/1). But since the edge traces of common films tested tend to be quite similar in shape, the gradient width alone serves as a good measure of film resolution. The rise time was measured from the 10 to 90 percent amplitude points of the edge gradient trace. The total rise time is defined as:

$$t = \frac{d + T}{s}$$

where, t = rise time from 0 to 100 percent amplitude points

d = spot image diameter on the transparency

T = width of edge gradient

s = linear speed of transition band spot across edge gradient.

This method was accurate to within ±35 percent from 2 to 20 lines per millimeter. Usable accuracy was limited to a maximum of 25 lines per millimeter due to the fixed spot size and spot sweep speed. The spot diameter employed was about .0005 inches. The equation above shows that the fixed values of spot size and sweep speed together swamp the contribution of edge gradient width toward the value of rise time. Optical glass limitations prevented a further reduction in spot size due to the spectral mismatch between the P16 phosphor (about 3700 Å peak response) and the microscope lens (about 5000 Å peak response). Sweep frequency of 1000 cycles per second could not be reduced without elaborate equipment modification.

Since the expected range of film resolution was from 5 to 40 lines per millimeter and time limitations would not permit a redesign, Method 1 was dropped.

The second method, spectral analysis, was intended to detect the smallest image detail. It was thought that the power of video frequency components generated by the flying spot might be related to the size of image detail for a constant sweep speed. A spectral frequency analyzer was used to display the video response as a function of frequency.

It was hoped that the response curve would take a sudden dip over a small frequency range corresponding to the smallest image detail. In other words, it was thought that the power of noise signal, above the highest image signal frequency, would be significantly less than the signal level of the smallest image detail. No sharp dips into noise were found. No characteristic shape correspondences were noticed between the video response curves and the associated film resolutions.

Apparently, the nature of the image-noise signal, along a scan line on the film, is to show a smooth and gradual transition in video response across the frequency range where the signal corresponding to smallest image detail falls off into noise; therefore, the second method was dropped.

The third electronic method was to vary the flying-spot scanner beam diameter and measure the corresponding changes in video frequency response. It was thought that the frequency response would not increase significantly for spot sizes below the diameter that is approximately equal to the smallest resolvable detail. This method, of course, could be vitiated by grain in some cases, where grain and smallest detail sizes are comparable. This method was dropped when it was learned that our equipment was to be ultimately limited to a minimum spot size of .0005 inches diameter. This corresponded to a maximum sensitivity of 25 lines per millimeter. The normal range of 5 to 40 lines per millimeter for this contract could not be covered without elaborate flying-spot scanner modifications; consequently, Method 3 was dropped.

The first and third methods might be successfully carried out, however, if a double-microscope, slit aperture, microdensitometer were used similar to those units commercially available from Ansco and Kodak. These systems produce minimum-sized light patches that are about .0001 inches wide.

The fourth method takes advantage of the linear relationship between useful limit of magnification and the film resolution in lines per millimeter. The relationship is expressed as:

$$R = LA$$

where R is the film resolution in lines per millimeter
L is the useful limit of magnification (pure number)
A is the acuity in lines per millimeter.

The useful limit is the point where a further increase in magnification would not produce an increase in photo image detail. In other words, the photo has been magnified just enough to show all image detail to the observer through the microscope. Further magnification would be useless. The acuity (A) refers to an average number of lines per millimeter seen on an unmagnified image at 10 to 12 inches viewing distance by the average observer under good lighting conditions. This works out to about 7 to 9 lines per millimeter at average image contrast for the average observer.

The technique proceeds in several simple steps. First, the operator would select the smallest image detail area on the photograph and verify it at the highest zoom magnification. A standard zoom microscope can provide continuously variable magnification from 3.5 X to 30 X. The operator might, say, select leaves of a tree as the reference detail in an aerial photograph. He would then reduce the magnification until the leaves begin to blur and lose their identity. The operator would then reverse the zoom control knob and increase the magnification until the tree leaf detail reappeared. At this time, the operator has bracketed the useful limit. Next, the zoom control is reversed for the second time and adjusted about halfway between the bracket settings. This setting is the estimated useful limit of magnification. If the setting is 4.5X, the film resolution is 4.5 times 8 lines per millimeter (average acuity value) or 36 lines per millimeter.

This method met with limited success over the entire range of film resolution tested. The chief disadvantages were the variability in individual acuity, estimates of useful limit, and operator fatigue.

In spite of the disadvantages noted above, the zoom method was the only method out of the first four methods covered above to give useful readings over the full range of film resolution expected in contract photography. This suggested that Honeywell should look for a better optical method in lieu of a better electronic method.

The fifth method investigated was the transition band (T-band) width concept. This goes back to the principles enunciated under Method 1 above. The difference is that the T-band width or edge gradient width is measured directly instead of being measured indirectly. The edge gradient profile is seen in Method 1, while the edge gradient's plan view is seen in Method 5. The fifth method was successful and was adopted as the standard means of measuring film resolution. This method employs a micrometer-eyepiece microscope to measure the width of sharp edge gradient or transition band between dark and light areas contiguous to the band. The relationship between T-band width and photographic resolution is defined as follows by empirical results:

$$\log R = \log a + b \log T$$

where R is film resolution in lines per millimeter

a,b are empirically determined constants

T is the transition band width in millimeters.

It is of interest to note that this relation is relatively unaffected by contrast variations above 2/1 contrast ratios, assuming that the photographs are not overexposed.

The micrometer-microscope method is superior to the other four methods investigated in accuracy, speed, and simplicity of operation. The range of operation will easily cover from 2 lines per millimeter to 100 lines per millimeter photographic resolution. One standard deviation is about ±11 percent and 90 percent of all readings will have less than ±20 percent deviation. The method takes only a few minutes per photograph and will work on both opaque and transparent photos. An operator can be trained to attain a satisfactory proficiency level in five working days. The equipment used consists of an American Optical student microscope and a Bausch and Lomb micrometer eyepiece. The magnification ranges from 50X to 1212X in four steps. The micrometer eyepiece can measure to within four millionths of an inch at 1212X. The equipment cost is approximately $400.

The measurement procedure starts with the subjective selection of several straight sharp edges in the photograph to be measured. Each edge is brought into the field of view and measured successively. The measurement of an edge consists of first aligning the eyepiece cross hair so that it is superimposed on either edge of the edge gradient transition band.

After setting the cross hair on one side of the T-band, the cross hair is moved across the T-band to its other side. The cross hair travel is read from the eyepiece micrometer knob. The resulting travel is entered on a graph showing the relationship between T-band width and film resolution. The film resolution corresponds to the coordinate value of the intersection of T-band width and the plotted curve. The resolutions, so determined, are arithmetically averaged to give the estimated photographic resolution.

The rule in setting the cross hair is to set it such that the tone of the area on one side is uniform up to the cross hair and the area on the other side up to the cross hair is not uniform. Obviously, this rule cannot be met in actual photography because of irregularities in the density distributions across sharp edges. However, the integrating capability of the observer's eye-brain allows him to average out irregularities by placing the cross hair so that it bisects the band of irregularities on each side of the transition band. The resulting cross hair placements appear to correspond approximately with the 10 and 90 percent ordinates of the density profile of the edge gradient.

The curve relating T-band width with photographic resolution was obtained from a set of 35 photographs ranging approximately from 2 to 100 lines per millimeter. Seven subjects estimated resolution of each photograph in lines per millimeter by identifying the Gurley bar chart pattern that was "just resolved." "Just resolved," in this case, refers to the subjective choice of the smallest set of six parallel bars (three are vertical and three are horizontal) in which the sense of the bar direction and number are preserved. The size of each bar pattern is inversely proportional to the number of lines per millimeter resolved on the film. Knowledge of the photo scale and pattern number is sufficient to calculate the film resolution from the Gurley Chart table.

The individual readings of seven subjects, based on the photographed Gurley Chart, were averaged for each photograph. These data were then plotted against the average values of T-band widths taken from several samples in each corresponding photograph. Four subjects read T-band widths with the micrometer microscope. These readings were averaged for plotting against estimated photographic resolution. The resulting scatter plot was used to fit the curve

$$\log R = \log a + b \log T$$

by the method of least squares.

The photo set consisted of Plus X, HyCon, and Commercial Ortho positive and negative transparencies. Contrast ratios varied from about 2 to 31. The film types and contrasts were not systematically varied relative to their resolutions. Some overlapping of film types occurred and showed no significant differences in T-band width as a function of film type at the resolution levels measured. This does not mean that all films will have the same T-band widths at the same photographic resolutions. It does mean that some films have such small differences in their spread functions that their edge traces will have nearly equal widths for a given film resolution. The results suggest that further research would be in order to establish other film group curves where the spread functions of the films differ significantly from those films tested.

The preceding comments do not refer to the effect of the camera lens spread function. This should be mentioned since the final spread function on the film is a function of the lens-film combination. Time and money limitations permitted a test of only two lenses. Some members of the set of 35 photos were taken with a 150-millimeter Schneider Symmar and the rest with a 47-millimeter Schneider Super-Angulon lens. No significant differences between lenses were found. That is, the scatter diagram, from which the T-band curve was derived, showed no deviation from a linear scatter distribution as a function of lens type. Therefore, lens spread function was considered unimportant relative to this study because no lenses other than those tested would be used.

A comment on exposure is necessary relative to the test set of photographs. It was found that overexposure would cause the bars in the bar charts in the photos to balloon. This ballooning, nicknamed the "fat-bar" effect, causes erroneous visual estimates of photographic resolution. The increased bar width reduces the spacing between the bars. The reduced spacing corresponds to the normal spacing of a smaller bar pattern properly exposed. In identifying the overexposed larger bar pattern as being "just resolved," the film resolution is estimated low for most of the lower contrast areas of the film which may not be as seriously overexposed. Since there is normally only a small percentage of very high contrast targets appearing, the readings taken with very high contrast bar charts have to be qualified when they show the "fat-bar" effect.

Contrast measurement did not require the development of a new method of measurement. The basic densitometric techniques currently in use were considered satisfactory. However, some special sampling techniques were developed to obtain precise measures of contrast in shadow areas that defined targets in the transparencies. The average contrast of a photo was too gross a measure because variations between the local and average contrasts ran as high as several hundred percent of the lowest of the two readings in some cases.

Honeywell had a Litton flying-spot scanner system capable of a 1000-lines-per-inch picture resolution, so it was decided to modify it into a microdensitometer. A cross-traveling slide stage and console were constructed. The unit was designed so that all measurements and adjustments could be made remotely. The output signal of the microdensitometer was displayed on an oscilloscope. A scope-mounted camera was used to record edge gradient profiles. Measurements of contrast were made as a function of relative amplitude between light and dark video signal levels which relate to the intensity of the light passing through the film as a function of flying-spot beam sweep position. The video response of the system was calibrated against a standard grey scale to within ±1 percent error.

PHASE II

Phase II of this contract breaks down into two task areas: procurement and grading of aerial and simulated aerial photographs.

The aerial photographs were procured from Rome Air Development Center, Photo Records Service Division, and Air Photo Surveys, Inc. Their subject matter consisted of simple tactical and strategic targets of military interest in about 35 target categories. Their resolution ranged from about 5 to 50 lines per millimeter. Scales ranged from approximately 1/5000 to 1/50000. The film was supplied as positive and negative transparencies on 5- and 9-inch formats.

Simulated aerial photography was made at Honeywell's Duarte, California facility. Two terrain model backgrounds with associated targets were made for this contract. Each terrain model is approximately 10 feet square. One model simulates an airport at a scale of 1/240, while the other model simulates a hilly forest area at a scale of 1/87. Scaled target models, such as aircraft and automobiles, were placed on the airport background. Weapons, land vehicles, troops, construction equipment and other targets were affixed to the hilly forest background.

The models were photographed outdoors at various sun angles and distances. The models were erected on stands so that their horizontal planes were vertical to the parking lot surface. This made it easy to photograph the models at varying distances to vary the ground resolution of the photographs. The photographic resolutions varied from 20 to 35 lines per millimeter. Scales ranged from approximately 1/5000 to 1/100000. All photographs of models were on 4 X 5 inch film.

The original condition of the photography, as supplied, had to be modified to suit the design requirements for image quality in Phase III. Specifically, each target had to be represented, in so far as possible, at each of 18 image quality levels. Ground resolution, for example, had to be represented in the following steps: 1-2, 2-4, 4-8, 8-16, 16-32, and 32-64 feet. Contrast ratio was to be covered in three steps: 1/1-3/1, 3/1-10/1, and 1/1-30/1. Since the original aerial photograph had 2 feet ground resolution or less, numerically, it was possible to degrade photos to higher numerical ground resolutions by variable defocus and multiple printing techniques. The model photography, on the other hand, was shot at appropriate distances at various lens apertures to get the desired ground resolution directly. The desired contrast levels were obtained by varying the exposure in making prints of the original transparencies.

The production of photographs for Phase III is still in progress. The results indicate that many of the smaller targets will not appear at higher resolutions and that some targets will not show maximum contrast desired.

The question of whether a photograph is acceptable or not is determined by its graded quality. If the photograph ground resolution and contrast simultaneously fall inside the ranges listed above, the photograph is acceptable. Marginal photographs are not being used.

The grading process involves the microdensitometer and micrometer microscope referred to previously. A reference graph of T-band width versus photographic resolution and a desk calculator are also used. If a processed photo is graded and found unacceptable, this fact is fed back to the processing lab to alter the processing of contrast and/or ground resolution.

At least one month and one week training are necessary, respectively, to operate the microdensitometer and micrometer microscope satisfactorily. Although subjectivity is involved in setting up and recording measurements with both instruments, reasonably standardized performance has been achieved.

Phase III will be covered by Mr. Meeker in the next presentation.

SIGNIFICANT DATA GATHERED

Not much significant data has been gathered to date. The Resource and Pilot experiments have shown, however, that the three most significant image quality factors are ground resolution, tone quality, and apparent magnification of the imagery. These experiments also revealed that the content factors, shape, target detail, and surroundings were the most significant cues in identification of simple, isolated targets appearing in the stimulus material.

It was also learned that a subject could respond to each and every circled target image in less than thirty seconds.

ANTICIPATED RESULTS

At this time five results are anticipated. I will simply list them as a matter of general interest and will not amplify the motivating arguments, which in some cases will be obvious. I will make an exception regarding the fifth anticipated result to clarify the statement.

1. Maximum contrast will not be as important as ground resolution in limiting photo interpreter performance. This implies that ground resolution will have to be controlled more closely than maximum contrast to produce black and white photos which are most efficiently analyzed by photo interpreters.

2. Photo interpreter accuracy and speed of identification will deteriorate significantly at lower contrast and higher ground resolutions. This implies that minimum levels of ground resolution and contrast must be specified in reconnaissance photography for efficient photo interpretation.

3. The selected targets will be ordered in accordance with the speeds and accuracies of each identification. This will imply comparative identifiability of targets studied and similar targets not studied.

4. There will be a close correlation between photo interpreter identification performance on aerial and simulated aerial photographs. This implies that future studies of photo interpreter performance can produce valid results by working with scale model targets and backgrounds.

5. Ground resolution will be found to be unsatisfactory as a good general indicator of the size of resolvable target detail. As an illustration of this, we noticed that a fighter aircraft approximately 45 feet long could not be seen when the photo containing it was degraded to 8 feet ground resolution. A check revealed that the tone areas defining the fighter were much smaller than the aircraft. In fact, the largest shadows were no more than 4 feet wide. The 4-foot widths should not be resolved at an 8-foot ground resolution as was found to be true. This shows that shadow size must be considered and that some sort of derating system will have to be applied to ground resolution which considers various factors, such as shadows. If this is done, ground resolution may become a more useful measure of resolved image detail in aerial photography.

PHASE III

PRELIMINARY CONSIDERATIONS

The given task called for in the original RFQ was to study speed, accuracy, and completeness of identification as a function of ground resolution. This task was interpreted as one of establishing norms of speed, accuracy, and completeness of identification as a function of resolution in order that the information obtained would have general applicability to the photo interpreter task. Based on this interpretation of the task, several preliminary guidelines were developed in delineating the approach to the problem.

To obtain applicable norms a wide range of ground resolution steps and a relatively large number of subjects would be required. To insure that individual differences in the experimental sample chosen did not confound responses made to different experimental conditions, it seemed necessary that each subject serve as his own control throughout all experimental conditions. The information obtained had to be applicable to a large range of target types. If our knowledge of form recognition were complete, the study of a small number of targets would permit extrapolation of the results to most of the tactical target types encountered in military photo interpretation. Since our lack of knowledge in this area makes it difficult to extrapolate the factors of form recognition from one form to another, it seemed necessary to study a wide range of target types so that the results might possibly be applied to similar targets not specifically studied. Target features had to be defined and the method of studying target features selected.

Other dimensions of the stimulus, in addition to resolution, that might be systematically varied or controlled were considered. Some of these were: Contrast ratio of target to background, scale of photos, blurring of image, photographic grain, time of day, context i.e., whether the target appeared in a background of rich surrounding detail, as against a limited surrounding, and use of photos of scale model targets rather than real targets. Not all of these variables could be adequately studied in a single experiment. Two were selected - contrast and model targets. Contrast was thought to be important because of the experimental evidence that the relationship between the form and background is perhaps the most important determinant of boundaries or edges and, hence, shapes. Use of scale model photography was thought to be important because of the possibility that future research around any of the previously mentioned variables could be more inexpensively and more effectively studied using models.

After selection of the contrast and model variables, an approximation of the size of the task was made. The RFQ called for testing at six or seven levels of ground resolution. A sample of 30 subjects seemed adequate to give moderately reliable norms. The target list, agreed upon by the customer, contained 30 to 35 target types. The target list was to be replicated with models across the three contrast levels and three resolution levels. The maximum number of individual identifications required would then be: 5 basic resolution levels X 3 contrast levels X 30 targets X 30 subjects, plus the three-fifths replication of models, plus about 5 targets at 3 contrast levels at a sixth level of ground resolution, or approximately 22,000 individual identifications. Since only two levels of any other factor would have doubled the total identifications required, the inclusion of other factors was essentially ruled out.

The number of identifications required meant that only a small amount of time could be spent on each identification. It appeared that a great time savings could be gained if the response measure of completeness were eliminated. The assessment of completeness necessarily requires some measure of search. Since search for some targets at the scales used might be a time-consuming task, it was decided to eliminate search, thus an assessment of completeness would not be made. There seemed to be several other good reasons for not including search: First, some time limit would have to be placed on search and any limit is arbitrary. "No search" is really a special case of placing a limit. Second, since we wanted to know the limits of the S's ability to identify the target when identification was the only task required, the "no-search" situation would give upper limits of accuracy and speed on identification. Third, including search makes it less likely that all subjects will respond at smaller photographic scales. In this situation, if the best norm is to be established, as many S's as possible should see the target. Fourth, to study the relationship of search and detection to identification, the two must be analyzed independently. In other words, search and identification are normally confounded in the interpretation process. To determine the proportion of time taken up in identification the two would somehow have to be separated.

There is one other requirement of the RFQ not already mentioned; this is the requirement to study identification of target features. Two questions arise: What are target features? What is the best way to study them? To determine what target features were important to the photo interpreter in identifying targets, a pilot study was conducted at the Army Electronic Proving Grounds, Fort Hauchuca, Arizona. A group of interpreters, ranging in experience from 5 to 20 years in the PI field, were asked to identify a number of targets at varying ground resolutions and contrast ratio levels. The subjects were also asked to indicate which features of the target were most important in analyzing their identifications. This was done in an open-ended type of interview. The target features most frequently mentioned in conjunction with identification across all levels of contrast and resolution were: Shape, size, shadow, surrounding detail, tone, and target detail. As to the second question of how best to study target features, the ideal approach seemed to be to manipulate the features of the target itself, either actually or photographically, and then correlate changes in the stimulus situations with changes in response. The shape of a target could be altered, for example, by adding nonfunctional areas or masses. On significant target details, e.g., the turret or tube of a tank could be altered or removed entirely. Such manipulation would be difficult to accomplish with real targets, but would be relatively simple with models. However, an equivalence between model and real targets has not yet been established. The tone, shadow, size, and other target features might be manipulated through photographic techniques of montage, masking, retouching, and the like, but such techniques would be extremely difficult to use with small targets at high numerical ground resolutions. Another approach to the study of target features would be to study tendencies to use the various target features in conjunction with identification as a function of resolution and contrast. This was the approach selected in the present experiment.

EXPERIMENTAL DESIGN

Based on the considerations reviewed above, the experimental design evolved as in Figure 1. The 18-cell matrix consists of three levels of contrast and six levels of resolution. Ideally, 30 real target types appear in each cell at all contrast levels across the

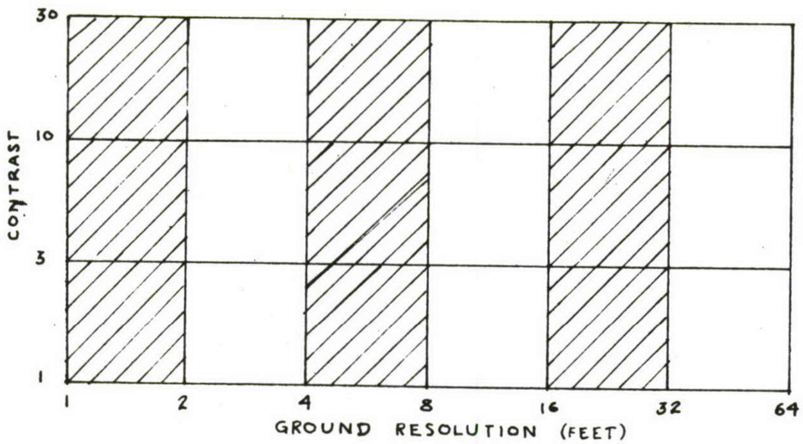


Figure 1. Experimental Design Matrix
(Shaded cells include both real and model targets).

first five resolution levels and 30 model target types (identical to the real target types) appear in each cell at all contrast levels across three of the five resolution levels. Since some of the target types selected tend to disappear at higher numerical ground resolutions, the number of target types in some cells is less than 30. At a sixth level of ground resolution, approximately 5 larger target types appear in each cell at all contrast levels.

EXPERIMENTAL METHOD

Stimulus Material

Thirty-five target types of tactical military interest were agreed upon, such as land vehicles, aircraft, field weapons, etc. Real photography was obtained through standard reconnaissance missions flown by the Air Force, supplemented by photography from PRSD files and from commercial sources. Model photography was provided by the contractor. To obtain the ranges of resolution and contrast called for by the experimental design, the photography obtained from these sources was processed in the laboratory. Positive transparencies were produced.

Subjects

The contract required that experienced interpreters be used as subjects. Because of military requirements, no large group of experienced interpreters was readily available for the extended time period implied in the design. An alternate approach was planned by RADC to use student photo interpreters currently in training at Sheppard AFB, Texas. The students participating will have completed all of their formal training on the type of targets used by the time the experiment is conducted.

Procedure

Photos with single targets designated to eliminate search are presented, one at a time, at the rate of one every 90 seconds. Thirty of the 90 seconds are allowed for responding and preparing for the next presentation. The 30 second-60 second time periods are based on information obtained in a pilot investigation conducted at Sheppard AFB, Texas, in which the procedures and apparatus to be used in the main experiment were tested.

Apparatus and Materials

The apparatus was designed to make possible presentation of a large number of targets to a large group in the 30 second-60 second time period described above. The apparatus can handle from 1 to 20 subjects at once. The apparatus consists of a cyclic timer in circuit with an Ektalight viewer and a strip chart event recorder. During each presentation, the timer starts the recorder and turns on the viewer. The light remains on for 30 seconds while the S studies the designated target. At the end of this 30-second period, the light in the viewer is automatically switched off and this event is recorded on the strip chart. If the subject is able to identify the target before the end of 30 seconds, he presses a button switch by the viewer; the viewer light is switched off and the event is recorded on the strip chart. Thus the speed of his attempted identification is recorded. The subject's identification of the target is indicated when he checks off his answer on a multiple choice answer sheet. The answer sheet is made up of a list of the targets which serve to decrease the probability of identification by chance. A list of target features also appears on the answer sheet. For each identification, the subject indicates any or all of those target features which he considers to have been most important in his identification. The recording of the target identification and features, as well as discarding the photo just studied and replacing it with another, takes place during the remaining time of the 90 second period. The answer sheet format will allow tab operators to transfer the response data directly to EDP punch cards. Data will be analyzed by computer, the basic statistical technique employed being analysis of variance.

GENERAL DISCUSSION

Dr. Roetling (Cornell Aero. Lab (CAL)):

You showed us some pictures where you were talking about edges and also mentioned lines. I wonder, in your analysis, whether you were actually using both thin lines and edges, or only edges?

Mr. Cook (Minn-Honeywell (M-H)):

The answer is "edges." It makes a difference, of course, if you use a thin line in which you may have overlap of the density distribution of parallel lines that are close together. This would alter the total amplitude of the edge trace at any one point on the "X" axis.

Dr. Roetling (CAL):

The next question, concerning your measurement of the width of the edge. I didn't see much real difference between the first method, in which you measured rise-time, and the last method, except that in the latter one you are relying on a personal judgment, whereas in the former you are using a spot scanner where I believe you used ten to ninety percent.

Mr. Cook (M-H):

You are exactly right. The basic idea is the same in both cases, that is, to measure the width of the edge gradient; however, the first method did have its limitation in regard to the fixed values of scanning speed and spot size. This means that for a given spot size we were limited to the number of lines per millimeter, which in this case was 25 lines per millimeter, as an upper limit. For the other method of using the microscope we were able to go up easily to a hundred lines; in fact the one curve I showed went from about two to a hundred lines. Despite throwing in another objective of a higher magnification lens, we were able to go to an increased magnification and therefore study lines that were close together, or in this case, edge gradients that would be narrower in width. The matter of studying an edge and placing the marker is truly a subjective process. However, when the observer is careful to line up the marker with his average edge position, (by observing a finite length of this edge) he can make a fairly accurate setting each time, which can be repeated. A man can actually realign the marker within about a ten percent tolerance band. It takes normally between two to five days of intensive training before he can do this. This points up one thing that becomes important as you increase the magnification, and may also indicate an upper limit toward the micrometer microscope method. When the field of vision becomes so narrow, due to higher magnification, your ability to integrate and place the edge accurately on the edge of the transition band deteriorates so that at some point, further up on the magnification spectrum, there will be such a small field a man will not be able to integrate on either the black or white side of the transition band.

Dr. Roetling (CAL):

If I understand what you are saying, then, if you had the electronic equipment that would achieve higher resolutions, you would probably use this rather than the human.

Mr. Cook (M-H):

That is true as far as the laboratory situation is concerned; however, the equipment we have costs about $30,000 and is fragile. Now if you are talking about a practical method for use in the field by interpreters, a .lying spot-scanner system is out of the question. So a well-developed training technique, in conjunction with this micrometer microscope, would be the best practical solution to a field-type measurement. On the other hand, the microdensitometer would certainly serve as a laboratory standard for checking out training procedures for the microscope. However, there are some severe problems that you are probably aware of. The existing spot sizes are much below a half-thousand, such as P-16 phosphor, which normally gives you a peak around 3700 Angstrom units which is considerably removed for the normal lens design area of around 5000 Angstroms (green light). Therefore, to work with a better lens system you would have to use a special ultraviolet lens (if you wanted to measure a P-16 phosphor). Of course, the P-16 phosphor is desirable because of its very rapid decay rate, and the closest one is more than an order of magnitude away in its decay.

Dr. Roetling (CAL):

I have one more question. You talked about contrast and rather amazed me when you said it was rather simple to define. Would you give a definition?

Mr. Cook (M-H):

We are talking about an intensity ratio. A ratio of about 30 to 1 would indicate that the energy measured on the brightest part of the scope would be about 30 times as great as that on the densest part.

Dr. Roetling (CAL):

Which is, then, simply using the brightest and darkest parts of the photo scan?

Mr. Cook (M-H):

That's right. For the contrast/measure used in this experiment we are talking about local contrast, where we are dealing with the areas of the shadows immediately defining the target itself. We are not talking about the average value of contrast over the entire field. This is essential because of the wide variation that you get from one local area to another and also because the PI will be studying just the area surrounding any one target, so that an average value would not be directly applicable.

Dr. Kraft (Boeing):

In line with the question of Dr. Roetling, I was rather interested in how you were distinguishing between the two methods for measuring edges. I was wondering if you had investigated the judged position of an edge as a product of the shape of the distribution on the shoulder of a curve. As you remember, several years ago, Vanderplass was very much interested in this estimation. And a second question (you will probably want to answer the two together): What is the criterion for accuracy of position that you used to evaluate the individual's performance to the one standard deviation for 11 percent of the width?

Mr. Cook (M-H):

The first question: We find that the lines per millimeter readings that we get by measuring the T-band are related to two things; not only to the width of the T-band but also to the shape of the curve. However, for the films that we have been working with, the shapes are enough alike that we can lump the readings across the three types of film and they all fall in that same distribution we were showing you. We find that by increasing the exposure, we shift the edge trace. This is the so-called "fat-bar" phenomenon, where, if you have a bar chart showing in the field and overexpose it (of course, you expand the bar in so doing), you find that measurement of the edge gradient remains the same although it shifts. This points out the fact that if one works with ordinary bar charts and is not careful to observe what his exposure is on the film, he can be misled as to his estimate of resolution. Answering the second question, dealing with the spread on the band shown, we have the reading of two men - an engineer and a lab technician. They each made a series of 35 readings, from which we concluded that, if the same two men had continued reading, 90 percent of their observations would have fallen within a 1.69 sigma band, where one sigma was 11 percent. In other words, 90 percent of the time either man, reading this type of photograph, would be expected to differ no more than 20 percent from the fitted average curve, in that distribution. Since we are talking about limited population of photographs, you might say that you cannot extrapolate too far. I should say this: the contrast range on these photographs varied from about 3 to 1 up to about 30 to 1. We are implying here that the contrast itself is not as influential as one might normally expect it to be. We found that the contrast plays only a significant role when you get down into the very low contrast regions. In other words, from about 5 to 1 up to 30 to 1, we find that the readings do not change substantially as a function of contrast.

Dr. Sinaiko (Institute for Defense Analyses) (IDA):

I wonder if you would comment a little about the cost of producing the models which you used and about the difficulty in making these. I was impressed by your high correlation between performance on the models and actual photographic images.

Mr. Meeker (M-H):

The correlation that I was speaking of is an anticipated correlation. At the present time, we have little data, if any, to indicate that the performance on the model photography is the same as the performance on the real photography. The indication that we do have is from student PI's at Sheppard AFB, who, without knowledge that these were model photographs (even though there were certain targets on the photographs that indicated these were models) gave no indication when questioned about it (said that they were not aware of any difference). These models measure 126 square inches, set on a movable screen, and come apart in two 5-foot by 10-foot sections. They were produced by Honeywell, at their display facility, by the Fine Arts Division. The basis of the photographic models that you saw here was polyethylene foam, covered with ground paint, and they are to scale. The targets themselves are of the type purchased in hobby and model shops. They vary in scale from the popular HO scale to 1/240 thousand. The models themselves, however, are very accurate and the background was made in accordance with models readily available. They are relatively expensive. Mr. Cook will tell you about this.

Mr. Cook (M-H):

Actually, we were rather proud that they did not cost as much as originally anticipated. The total cost of the two targets with models, we put at about $5200. I would expect that if you wanted to do them outside your own shop the cost would be about $10,000.

Mr. Meeker (M-H):

The importance here too is also that the main cost was the labor, as skilled technicians were required. The cost of the material was relatively small.

Mr. Speer (Houston-Fearless):

In talking about your contrast ratio, were you referring to the original scene contrast, or to the recorded film record of contrast?

Mr. Cook (M-H):

The answer to that is the second case.

Mr. Speer (Houston-Fearless):

Then actually your scene contrast (since film is nonlinear) might have been up to 1000 to 1 or more.

Mr. Cook (M-H):

Yes, that is a possibility, especially in the area of the high contrast targets.

Mr. Speer: (Houston-Fearless):

Then you commented on the field scanner, and its portability and ruggedness; did you ever consider anything other than an electronic scanner, such as a simple moving mirror, of the type we use in our image quality meter to create a moving spot?

Mr. Meeker (Houston-Fearless):

Yes, we have seen your approach there and it looks like a very good one. We happened to have the flying spot scanner system from a previous program and that is why we used this system. However, in the future it is entirely possible that a system such as Houston-Fearless has used with a moving mirror will be desirable.

QUALITY CATEGORIZATION OF AERIAL PHOTOGRAPHY

Dr. Paul G. Roetling
Cornell Aeronautical Laboratory, Inc.
Buffalo, New York

ABSTRACT

A study is in progress at the Cornell Aeronautical Laboratory to determine a "summary measure" of image quality. The measure will relate physically measurable quantities in the image to photo interpreter performance. A program is described in which real aerial photography is selected and photographically modified to achieve a range of "qualities." These modified photographs are then used in a series of tests in which photo interpreters perform realistic interpretation tasks. The results of the tests will be used to derive a summary measure of image quality.

INTRODUCTION

The study and the experimental program discussed herein are being conducted at the Cornell Aeronautical Laboratory under Contract AF 30(602)-2684 with the Rome Air Development Center. The program is directed toward the derivation of a summary measure of image quality for photo interpretation.

It is now possible to formulate equations which relate the photographic image produced by a reconnaissance system to the original object scene. These equations contain parameters describing the effects of atmosphere, lens, film and image motion on the image. Although the values of several of these parameters still require some investigation, and are in fact being investigated on other contracts, the equations can be written in general form. However, there is at present no equation which relates the final photographic image to the potential performance of a photo interpreter in extracting information from the image. This missing link is the subject of the current effort.

If such an equation can be derived, that is a measure of interpretability expressed as a function of the physical image characteristics, several uses are apparent. The summary measure of image quality would provide a rating scale with which priorities could be established at times when the image acquisition process overloads the interpretation process. Further, by examination of the components of the quality measure, it would be possible to predict those cases in which enhancement is possible and foretell the form of enhancement which would be of the greatest value in increasing the quality. Using this approach, it can be seen that the summary measure also provides a design criterion for the synthesis of new reconnaissance systems. Lacking the summary measure, system optimization has been limited to a somewhat intuitive process.

This program has been designed around a basic philosophy of remaining close to realism. Only real aerial scenes, not abstract shapes or models, will be used. Trained photo interpreters will be tested in the interpretation tests, and only realistic interpreter tasks will be used. In order to constrain the program to reasonable test size and keep within financial limits, several limitations were imposed. Only black and white photography will be considered and only monocular viewing is included. Both color and stereo add considerably to the complexity of the physical description. The physical parameters used will not be a complete description of the image, but rather are a simplified description which is both comparatively easy to handle and precise enough to treat interesting situations. Finally, the ranges of variables are essentially limited to those which might occur in present day aerial photography.

The general approach to the problem is shown in Figure 1. Aerial photographs are selected according to a set of requirements on target content and scale. This imagery is then modified with respect to the physical quality parameters, creating a set of photographs in which the physical quality varies over the chosen range. The physical parameter values of each photograph are then measured, and the photographs are used in a series of interpretation tests. The performance scores of the photo interpreters are then correlated with the physical parameter values to obtain a combination of physical measurements which increases monotonically with increasing performance. Each of the steps of this approach are now treated in more detail.

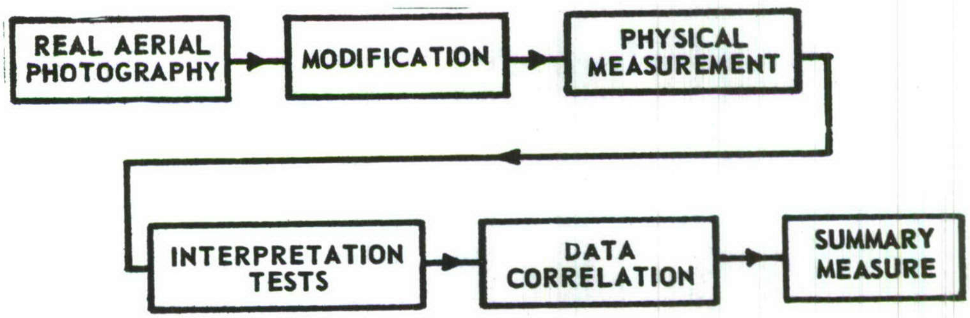


Figure 1. Approach to Problem.

INPUT PHOTOGRAPHY

Target Specifications

Each original image selected for test purposes must contain a minimum of ten unique military targets. In this sense, a component of a large target area may be used as a target (i.e., a missile might be one target and an oxidizer truck another, both within one missile complex). In order to provide near-operational characteristics, only normal military targets would be chosen. A range of photo scales would be used with a corresponding range from tactical to strategic interpretation tasks.

Each target image must be positively identified on the original, either by ground truth or by concurrence among several trained PI's. The targets must be identifiable on the highest quality images shown in the tests, and must vary in difficulty of identification. These requirements assure that every target contributes significant test data and that interpreter scores vary smoothly with variations in image quality, allowing the complete quality range to be measured.

Physical Image Description

Other than the scale, which simply sets the image size of the photograph, three quantities must be described. These are the "contrast," describing the magnitude of the density fluctuations in the photograph, the "smear," describing the fidelity with which each object point has been reproduced in the image, and the "noise," describing the random fluctuations in the image which are not a part of the object scene. These three quantities are considered in order.

Contrast

For a simple target (i.e., uniform brightness, B_T, on a uniform background, B_G) the contrast is easily defined. Although several mathematical forms are commonly used, the simplest is:

$$C = \frac{B_T}{B_G} .$$

The logarithm of this contrast is frequently used, and is convenient because the density difference between the target and background on a photograph is then:

$$\Delta D = D_T - D_B = \gamma \log C$$

where γ is the film gamma. Thus, ΔD is often called the "contrast."

Since a photograph consists of many complex targets, where, in fact, the term target is not predefined as any object may be a target, the definition of a total image contrast becomes a problem. A simple extension might be to consider ΔD as the difference between the greatest and least densities in the photograph. While this definition is sometimes used, a different extension has been chosen for this project. The density difference between nearby points is used. In order to get a measure which is not sensitive to a single bright spot or dark spot, a number of the largest density fluctuations* are determined and averaged.

Smear

The fidelity with which each object point is reproduced in the image can be described by the intensity distribution of the image of one point, and is commonly called the point spread function. It is now common practice in optics and photography to use the Fourier transform of the point spread, which is called the sine wave response. This function can be measured for a system by photographing a set of test charts which have an intensity which varies sinusoidally with position, and in which the size of one cycle of the sine wave is different for each chart. The ratio of the contrast modulation (brightness difference divided by average brightness) of the image to that of the object is plotted as a function of spatial frequency (number of sine wave cycles per millimeter). The sine wave response of a typical system is shown in Figure 2.

Since sine wave test charts do not commonly occur in aerial photography, whereas edges do, the mathematical relation of the point spread to the imaging of an edge is used

* A "fluctuation" is defined here as a monotonic increase or decrease.

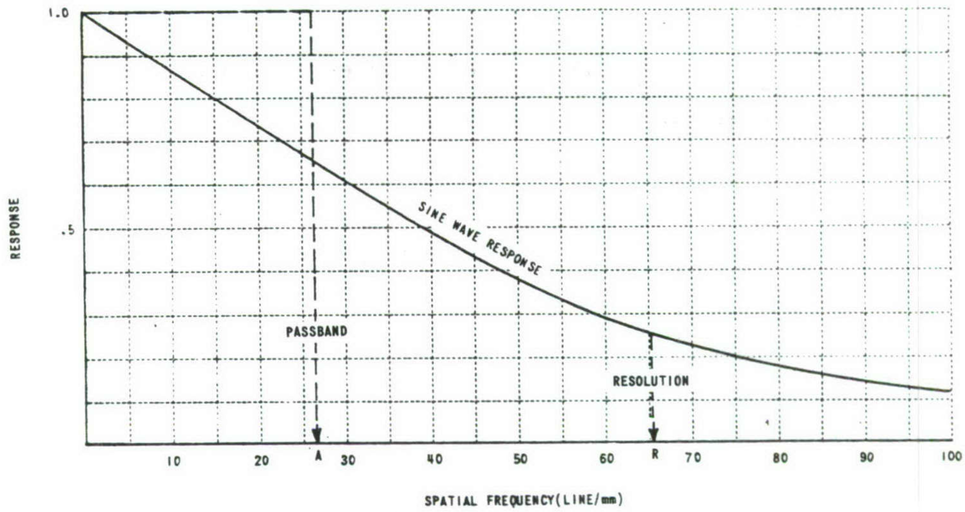


Figure 2. Typical Sine Wave Response.

to obtain the sine wave response in practice, as described in the measurement of variables.

Consider now measures which can be used to describe the information given by the sine wave response function.

In general, the sine wave response function could be a function of orientation; that is, the curve shown in Figure 2 could change if the sine wave charts were rotated. While this variation occurs far off-axis for most cameras, for reason of simplicity it is ignored at present and interest concentrated on near axis portions of photography.

Two measures are currently being used on this project to describe the sine wave response function. These are: first, the area under the square of the sine wave response curve, called the passband, and second, the 25 percent point of the curve, which for lack of a better term is called resolution in this paper. The rectangle shown in Figure 2 is that rectangle which has the same passband as the curve shown, so point *A* is the passband value. Point *R* is the resolution value.

Some insight to the meaning of these variables can be gained by further examination. The passband, originally suggested by Schade,[1] is closely related to acutance[2,3] and correlates with the visual impression of sharpness. The resolution can be shown to be essentially the resolution value which would be measured with a low contrast (1.4:1) test chart imaged by a normal camera lens on a very low grain emulsion. Thus, the 25 percent point value is not the usual high contrast (and usually grain-limited) resolution, but rather is a measure of detail reproduction at normal contrast (limited by the eye).

Noise

The noise in an aerial photograph is formed by the granular nature of the emulsion. This noise is a function of the emulsion used, the development process, the average density level and the spatial frequencies over which it is measured. The first two factors determine the results obtained, but need not be of concern in measuring the granularity.

The third factor can be taken into account most readily by making the granularity measurement at the average image density and ignoring the slight changes with density, since in all cases the grain variation with density is similar (less at lower densities). While the grain noise in copied or enhanced photos can be a function of spatial frequency, original photography has a nearly flat grain spectrum. Thus, the grain noise can be measured in a manner similar to that used by Eastman Kodak (results shown in film data books), measuring the root-mean-square density fluctuation in a "uniform exposure" region.

Treatment of Scale

Scale has not been mentioned as a physical variable, and deserves special attention. When scale is discussed, confusion often arises since in reality two different scales are referred to. The block diagram in Figure 3 is an attempt to clarify the situation. When an aerial photograph is acquired, the image size is related to the scene by the "photographic scale." When the image is to be interpreted, the image is normally increased in size for viewing, either by photographic enlargement or by optical magnifiers (or a combination). This latter magnification has been called the "viewing scale" for convenience.

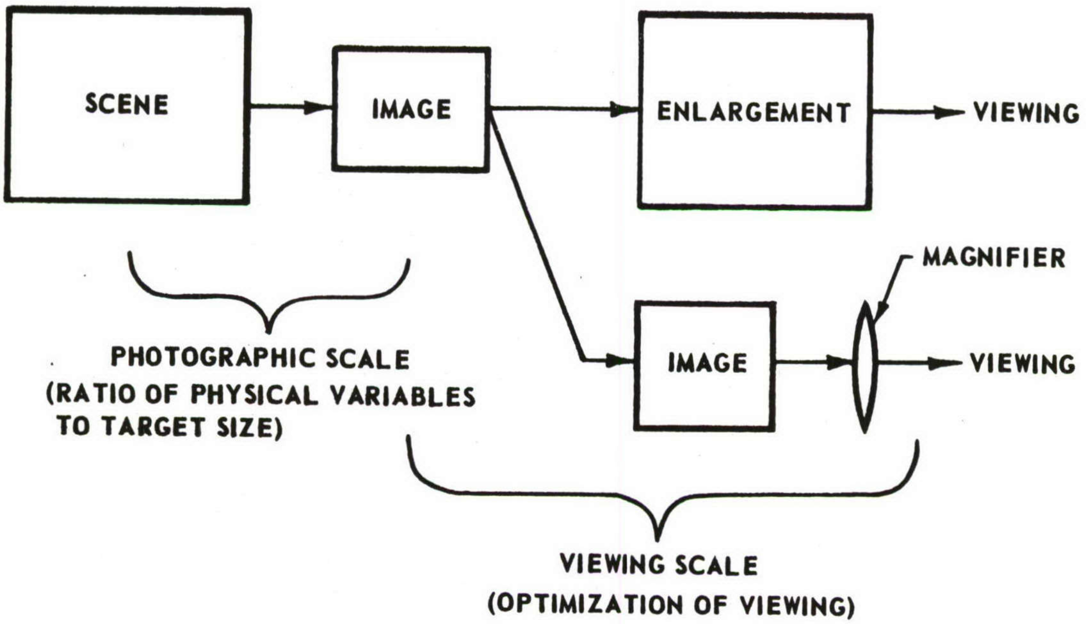


Figure 3. Treatment of Scale.

Considering the viewing scale first, the purpose of the change of size is to present an image to the eye under optimum conditions. Clearly, the amount of magnification which is useful depends on both the target to be examined and the quality of the original image. The photo interpreter is the best judge of optimum conditions for each case, and is therefore allowed to vary the viewing scale by selection from a set of magnifiers.

The photographic scale sets the ratio of size of a given target to the dimensions involved in the physical quality measures. Thus, for a given quality, the photographic scale controls the target types which might be interpreted. Minneapolis-Honeywell is currently studying this relation in terms of ground resolution. In order to study quality, independent of this target-type effect, the targets are selected on each original as being appropriate for the particular photographic scale. As already stated, photographs at various scales are selected. After selection, every time a given original is presented the photographic scale is held constant.

Selection of Variable Values

The variables just described are not completely independent. In order to perform tests with samples of photographs which are representative of the distribution which exists in real photography, the ranges of the individual variables and their correlations must be taken into account.

Contrast is relatively independent of the other variables. A correlation with scale might be noted since higher altitude implies lower contrast and generally smaller scale. This correlation is weakened by the use of various focal length lenses (which decorrelates altitude and scale) and by natural weather variations (which introduces contrast change at one altitude). Therefore contrast will be assumed to be independent of other variables. The range of contrast is taken to cover from barely detectable contrast to as high as normally observed in aerial photographs. In average density fluctuation measures of contrast this limits the range of ΔD to approximately : $0.05 < \Delta D \leq 1.0$.

In order to determine representative ranges and correlations of the passband resolution, and grain variables, a number of lens-film combinations were considered with lenses at several focus settings (coarsely simulating normal focus errors and other errors, such as motion or vibration). The resultant values are plotted in Figure 4 with theoretical limits shown. Although the cases shown are by no means complete, ranges and correlations can be seen.

The granularity σ value range is: $0.01 < \sigma < 0.15$ and is typified by several common films. The correlation with passband and resolution is essentially one of setting an upper limit depending on σ. In practice, a low grain film would seldom be used with a poor camera system, thus tending to eliminate low passband and resolution at low σ. This cutoff is not shown, but would be included in sampling.

The correlation between passband (A) and resolution (R) shows that the relation is essentially linear, or: $A = KR$ where the value of K can vary in the range: $0.25 < K < 0.60$, independent of the value of R. While R is limited only by current lens design, a practical limit can be established for the test photography. Since the resolution limit of the eye is the order of 5 to 10 lines/mm at normal viewing distance, the interpreter cannot make use of fine detail unless magnification is provided. Thus, the upper limit of the resolution used is chosen after the maximum available magnification is determined. Up to 12X magnifiers will be provided and the value of R will range: $5 \leq R \leq 50$ lines/mm, thus covering a range from degradations nearly visible by eye to those visible only under maximum magnification (a 4 line/mm limit is chosen at the high end to assure that sufficient magnification is available).

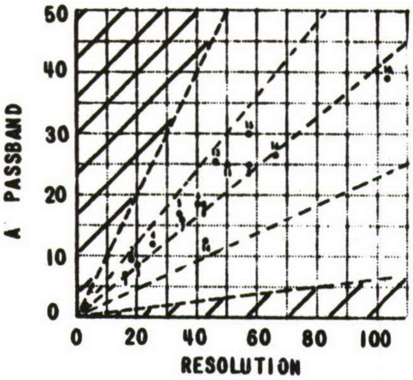


Figure 4a PASSBAND - RESOLUTION

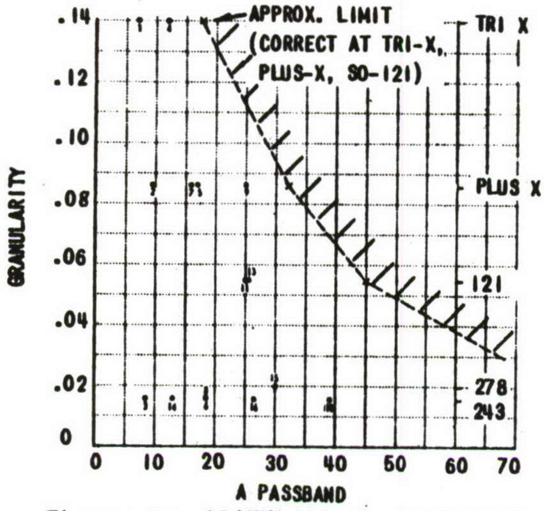


Figure 4b GRANULARITY - PASSBAND

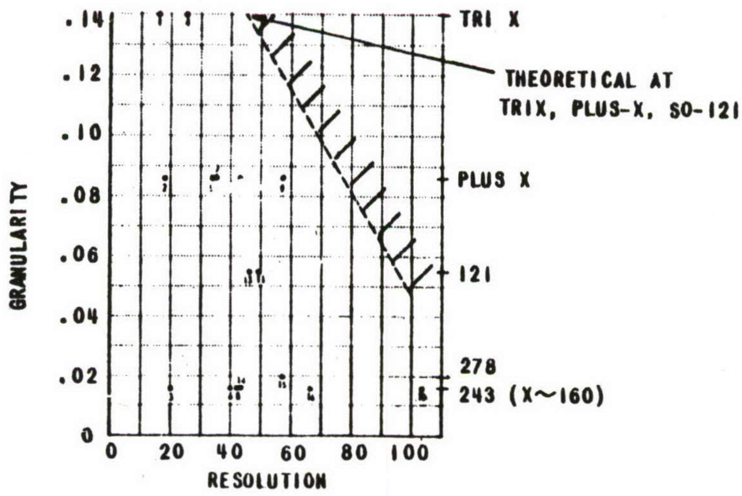


Figure 4c GRANULARITY - RESOLUTION

Figure 4. Variable Correlations.

IMAGE MODIFICATION

A number of methods are available for image modification. Table I lists several of these methods, what each primarily modifies, and whether the method enhances or degrades that parameter. Generally, side effects (for example, method 4 also changes granularity slightly) are also present and must be taken into account, but these will not be considered now.

TABLE I
MODIFICATION TECHNIQUES

METHOD	MODIFIES	ENHANCE OR DEGRADE
1. DEFOCUS	τ	D
2. COPY ON GRAINY FILM	$G, (\tau)$	D
3. COPY WITH VEILING ILLUMINATION	C	D
4. COPY ON HIGH GAMMA FILM	C	E
5. SPATIAL FILTERING	τ, G	E OR D

The last method, spatial filtering, is probably least familiar, therefore a diagram of the system is shown in Figure 5 and a photograph of the system is presented in Figure 6. The original photograph is placed in the object holder. The spatial frequency spectrum of the scene appears in the filter plane, and the modified image appears in the image plane.

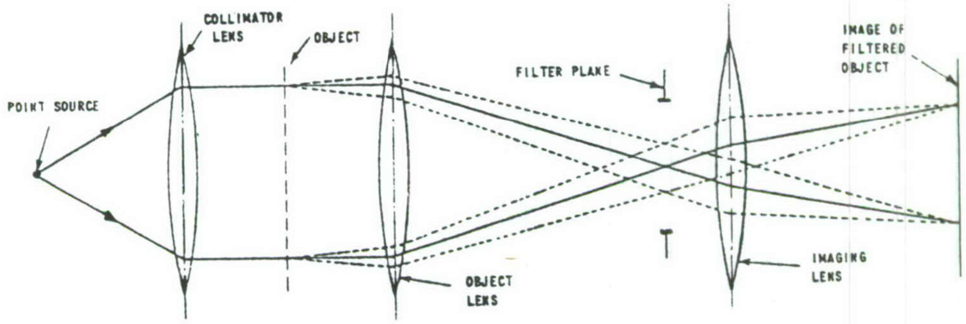


Figure 5. Three-Lens Spatial Filtering System.

The passband and the resolution are increased or decreased by changing the spatial frequency response by placing appropriate filters in the filter plane. The resultant spatial frequency response is the product of the original response and the filter function.

Granularity is decreased by filtering out any frequencies which are not needed to form the filtered image. Increased granularity is obtained by imaging the modified scene onto different films (2 in table).

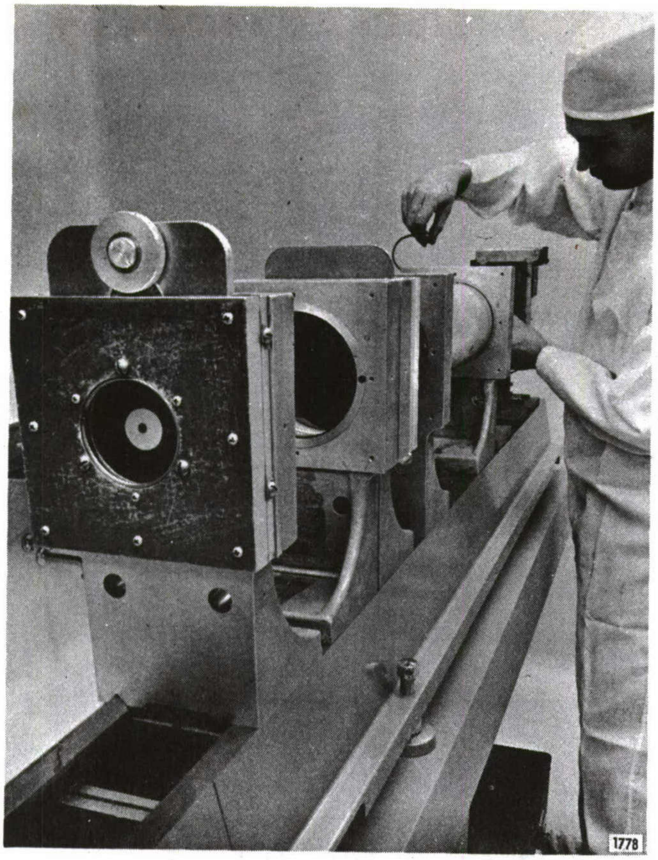


Figure 6. Spatial Filtering Equipment.

Contrast is modified by two means. Increased contrast is obtained by developing the image to a gamma greater than unity. Decreased contrast is obtained by dividing the total exposure time into a combination of a short uniform veiling exposure and the remainder as an imaging exposure.

MEASUREMENT TECHNIQUES

All measurements will be made by the use of the CAL-owned Ansco microdensitometer. Three scans are required, one to measure contrast, a second for passband and resolution, and the third for granularity. Figure 7 shows a photograph and the three typical scans used.

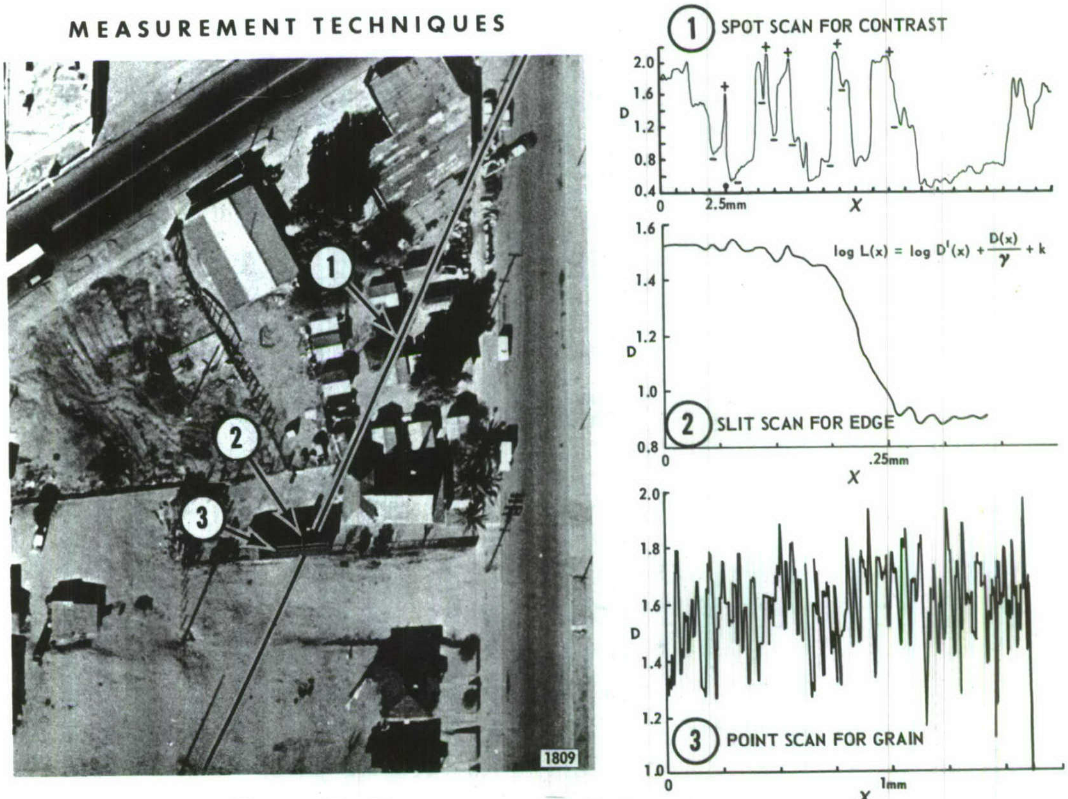


Figure 7. Measurement Techniques.

Contrast Scan

The contrast scan is a long scan, across most of the image, chosen across a representative portion of the scene. A circular aperture is used which is large enough to integrate small fluctuations such as grain. The trace (1, in Figure 7) is then examined for large monotonic density changes (some are marked on the figure). A set (usually picked as 10) of the largest changes are then averaged to determine contrast.

Edge Trace

The passband and resolution are determined from an edge trace. An edge is selected which is known to have been a sharp transition in the original scene. For example, the top of a peaked roof (shown as 2) is a different brightness on each side of the peak. If the top were rounded, this would not be a sharp edge originally and would not be used. Since judgment is required, a single edge should not be trusted, but rather several are used. A slit aperture is used to integrate over grain and record the edge without degradation.

The edge trace obtained is related to the line spread (point spread integrated in one dimension) by the expression shown on the figure. If the value of γ is unknown, two edges are used and an iterative procedure yields the best value. K is only a normalizing constant. In order to arrive at a good value for $L(x)$, the data are smoothed by a computer program to eliminate noise. Once $L(x)$ is obtained, its Fourier transform is computed to yield the sine wave response, from which the passband and resolution are computed.

Granularity

The granularity is measured in any "uniform" area (such as the rooftop shown as 3), by scanning with a small spot. The standard aperture is a 24-micron-diameter circle, but it is convenient to use a 10-micron circle and correct the data for aperture. The area is scanned, providing a trace as shown in the figure, and the data are simultaneously sampled, converted to a digital form and punched on tape for direct input to a computer. At least 1000 sample points are collected. These data are used to compute the average density and the root-mean-square density fluctuation (σ). Depending on the aperture used in scanning, the measured σ is converted to a standard (for 24-micron aperture) by:

$$\sigma_s = \sigma \left(\frac{D}{24}\right)$$

where σ_s is the standard value and D is the diameter of the aperture used (measured in microns).

TEST PROGRAM

The test program will be conducted in three phases. Phase I consists of pretests which will serve to clarify doubtful points about the main test program and remove any test difficulties. Phase II is the main or determination test which will be used to derive the summary measure of image quality. The last tests, Phase III, will be a validation of the summary measure.

Pretests

The pretests serve a threefold purpose. Like all pretests, they provide an opportunity to check and refine test procedures. They will provide a preliminary estimate of relations between performance and physical quality. For the program, the pretests will also provide a valuable comparison of interpreter tasks.

Perhaps the most realistic photo interpretation task that could be used is "free search." For this task, the interpreter searches the photograph and locates and identifies any targets he finds. Unfortunately, since the total number of responses is not specified in any instruction, the interpreter is left to choose for himself whether or not to respond to doubtful targets. This creates a conflict between accuracy and completeness. To increase accuracy, one should respond only when he is certain, while to increase completeness one should respond to everything. Such a set of conflicting criteria will result in an increased error variance and masking of the true effects of experimental variables. Therefore, other similar tasks, but without this conflict, will be used.

Two tasks will be tested during pretests, "identification" and "modified free search." Each task provides a specification of the number of responses and establishes an unambiguous accuracy criterion. For "identification," the targets to be identified are annotated on the photograph and the interpreter identifies each target, choosing his answers from a list which contains these target identifications plus many target names which do not occur in the scene. This task requires a minimum of interpreter time. In the modified free search task, the interpreter is instructed exactly as in free search, except that the total number of responses is specified. In this case, the interpreter responds to the most certain target first and continues until the correct total number is reached. This task is somewhat more realistic than identification, since the photographs are not annotated, but the task requires more time. These two tasks will be compared during the pretests.

The pretest results, then, will provide a comparison of tasks and an initial estimate of the summary measure. Scattergrams will be plotted to show the shape of the relation between performance and each variable. Since the multiple regression analysis to be used is linear with respect to the variables used, these scattergrams will allow other combinations of functions of variables to be chosen which vary more nearly linearly with performance.

Determination Tests

The determination tests are the main tests, used to derive the summary measure of image quality. A reasonably large sample of photographs and interpreters will be used. The task to be used in the determination tests will be selected on the basis of the pretests. The identification task is preferred for simplicity and total testing time. If significantly different results occur in the pretests for the two tasks, a decision will have to be made as to which task is more representative of the tasks for which the quality measure is to be used.

As in the pretests, multiple regression analysis will be used to derive the summary measure. The combinations and functions of variables included in the analysis will be those determined from the pretests and from further examination of functional relations indicated in the determination tests.

Validation Tests

As a test of the general validity of the summary measure, a final small test will be performed. A representative sample of photography will be chosen and the physical quality measured. The summary measure will be used to predict the performance on these photographs. Tests similar to the determination tests will be run and the observed performance will be compared to the predictions from the summary measure.

CONCLUSIONS

The program described is an attempt to derive a measure of interpretability in a series of realistic tests. The measure of quality is recognized as a combination of the effects of several physical variables. The resulting measure will be assumed to be successful if it can be shown that a monotonic relation exists between the value of the summary measure and interpreter performance.

REFERENCES

1. Schade, O.H., *JSMPTE*, Vol 64, p 593, 1955 and previous papers in series.
2. Higgins, G.C., and Jones, L.A., *JSMPTE*, Vol 58, p 277, 1952.
3. Perrin, F.H., *JSMPTE*, Vol 69, p 151, 1960.

GENERAL DISCUSSION

Mr. Farina (RCA):

Looking forward to the time when your summary measure is used in the operational setting, can you give an estimate how long it might take to measure any one photograph? Or would this be done by machine?

Dr. Roetling (CAL):

It would take us quite a while right now to measure any one photograph. It might be possible for the people of Houston-Fearless to do this very rapidly. They are working on something called the "image quality meter" that measures these same parameters. I think the question of how long it takes to measure the parameters of a photograph would be better directed to someone who is working on similar instruments, rather than someone who is trying to find out, first of all, whether the relation exists, and, secondly, what it is if it does exist. Our measurements are being made on a microdensitometer and are necessarily very slow. They do have the advantage, however, of being precise. I do not know whether someone from Houston-Fearless will try to compare the microdensitometer with their instrument or whether others may be working on similar instruments of which I am not aware.

Capt. Millspaugh (ATC):

I have three questions I would like to ask you, Doctor. One is in terms of the material you are using. Are you going to conduct your tests primarily with photographs or have you considered using positive transparencies?

Dr. Roetling (CAL):

I should have clarified this. These are only positive transparencies. The normal photographic paper simply does not have the quality of transparency materials. We would like to obtain higher qualities than the paper will handle and have essentially agreed that interpreters are more familiar with positives than negatives. We have decided on that basis to use positive transparencies.

Capt. Millspaugh (ATC):

The second area on which I was trying to get some background was the group you are using. You mentioned you were going to use experienced PI's. Could you identify that group?

Dr. MacLeod (RADC):

We are going to conduct our pilot tests with the group at the Army Personnel Research Office. From that point on we hope to make use of some of your people, assuming their availability.

Capt. Millspaugh (ATC):

The third question I would like to ask you has been presented to the gentlemen from Minneapolis-Honeywell prior to this meeting. I would like to find out the ground rules you are using in the tests you are giving the PI's. Are they going to use primarily just magnification instruments or are they going to scale the photography, make measurements, or use reference materials such as PI keys and recognition manuals?

Dr. Roetling (CAL):

As far as the first part of your question is concerned, we are planning on providing non-stereo magnifiers which have been tested for their effect on image quality. I think we will be necessarily extending the magnification range beyond what the interpreters probably are trained with but are hoping to stay below, say, about 10 or 12 power. In a pretest, we are required to make some kind of simple measurement to locate a target. This would probably be a grid mark on the photograph. We find it very convenient to use the Xerox process for marking the photographs. We will probably mark a grid on at least the sides of the photograph and ask for some simple measurement for the location of a target.

Capt. Millspaugh (ATC):

The reason I brought this question up is that (speaking as a PI) no PI will put his confidence in a report unless he can work out what the image is in detail. And this refers to the fact that he must make measurements in the case of identifying small objects, for example, an aircraft. There are a great number of aircraft that have the same configuration or shape. For him to identify an aircraft, he would have to make linear measurements and refer to other aids (PI keys or recognition manuals). Also, he consults with other individuals, because no work of a PI really in essence is the work of one man; it is the result of group consultation. So this just makes me question the validity of the responses that the individual can give if it is circumscribed in terms what he will do or what he will use in making the interpretation.

Dr. Roetling (CAL):

I would say that the point on the consultation would magnify the problems in our test; we probably wouldn't be able to allow this. If it becomes necessary to refer to PI keys, etc., I see no reason why this could not be allowed. We would, in fact, like the method to be as near to the PI task as possible, and, before we go in this program much further, I would like to talk to you about this.

Mr. Hauser (Ft. Huachuca):

For sake of a more comprehensive series of tests, why have you not included stereoscopic analysis and why have you not included skills and backgrounds as part of your tests? I think that, as a part of the validity of any test, hinges heavily on these questions.

Dr. Roetling (CAL):

I will try to answer your questions item by item. Inclusion of stereo is partly a question of getting stereo coverage. Also, if you look at a picture in stereo, you get a grain effect that is different from looking at a single photograph. We hope to just tackle the single photograph now, and then at a later time try other things that you have listed. I did not mean to rule them out forever, but only temporarily for the sake of the size of the test program. The last point of skill and background is simply a matter of repeating the tests with different levels of skilled interpreters. I think that again it is a matter of size of the program.

Dr. MacLeod (RADC):

I think that both of these questions are good ones, and we would be the first ones to admit that our studies are certainly not complete in terms of all of the elements that could have been introduced; but we feel that at least we have made a good beginning, and certainly want to keep these questions in mind in any follow-up program. We obviously would be concerned with extending these programs to include stereo and other modes of response and in validating our studies by seeing if we tend to get the same kind of results when we move into other situations which give us a greater amount of realism. We have to make some degree of compromise, obviously, in terms of the amounts of funds and time that we have, and we are trying to select parameters in the experiment that are not too far removed from reality.

Mr. Sorem (Eastman-Kodak):

We are really gratified with the realization that transparencies can produce aerial photographs of so much better quality. We have advocated for a long time that, if you want to get the most out of aerial photographs, you should use positive transparencies rather than paper prints, for several very obvious reasons. I think the main obstacle is that PI's almost unanimously insist on stereoscopic viewing. Manipulation of transparencies for ease of stereoscopic viewing presents a few relatively minor problems which can be overcome. This is the main thing, aside from the relative cost and the fact that you want paper prints also for wider distribution. However, I think that as time goes on there is no question but more and more transparency materials will be used.

Dr. Roetling (CAL):

I could throw a question out here which might be more appropriately directed to the people from Sheppard; I think that Mr. Sorem would agree with me that you are not only going to stay with transparencies but, it you really want quality, you will also want the original; and therefore, you should not really be using the transparencies, but should be using the original negative. So long as silver-halide is a negative working material, I think you want to keep down the number of generations that you go through and it would be nice if we could train people to be familiar with white shadows. As I said, we do want to work with transparencies; interpreters have been trained to work with positives, so we will stick to positive transparencies. However, I feel that in the future, we will probably want to include negatives.

PSYCHOPHYSICAL ASPECTS OF IMAGE QUALITY

Dr. Robert Sadacca
Dr. Alfred I. Schwartz
US Army Personnel Research Office
Washington D.C.

ABSTRACT

The Army Personnel Research Office is sponsoring a study to evaluate relationships between psychophysical judgments of image quality and physical image variables. Relationships will also be established between their judgments and actual interpreter performance. The general hypothesis is that viewers are able to judge efficiently the overall quality of imagery and thereby shortcut the need in the Army tactical setting for measurement devices and the application of weighting formulas in assessing image quality.

Recently, The Department of the Army made available to the Army Personnel Research Office (APRO) additional funds to be used in the study of image quality. It is planned to contract out these funds for a study of the psychophysical aspects of image quality with the ultimate aim of setting up psychophysical scales by means of which image quality can be judged directly by a human observer. This research will be planned so as to complement the research currently being undertaken for the Air Force by the Cornell Aeronautical Laboratories and Minneapolis-Honeywell. These organizations are exploring the relationships between physical image variables and performance. The APRO sponsored study will deal with the relationships between psychophysical judgments of image quality and physical image variables. Also we want to study the relationships between these judgments and the actual performance of interpreters examining or using the imagery. The general hypothesis is that humans are able to judge efficiently the overall quality of imagery and thereby shortcut the need in the Army tactical setting for measurement devices and the application of weighting formulas in assessing image quality.

A possible end-product envisaged for this research is a set of images forming one or more psychophysical scales. Suppose, for example, it has been empirically determined that there are three psychophysical dimensions in judgments of similarity of images in regard to their quality. Let us say these dimensions can be identified with, or related to, the physical dimensions of, photo scale, sharpness, and contrast. Then, the psychophysical scales could be arranged in book form with each page in the book containing a set of images varying systematically in regard to sharpness and contrast but with a fixed photo scale. Underneath each image on the page, the values of the image along psychophysical, physical and performance dimensions could be specified.

When a measurement of image quality is needed, the image interpreter or lab technician could very rapidly determine the photo scale of the image that needs to be measured. He would turn to the appropriate page in the book, compare the test image with the standard images on the page and read off the performance or physical values associated with the standard image judged to be most similar in terms of quality to the test image.

A public announcement concerning the general terms of the contract to be let by the Army and monitored by APRO will be made in the near future. In order to provide potential contractors with some empirical information upon which to base their proposals, APRO has conducted two small probes in this area. The APRO research effort was limited by lack of sophisticated photolab facilities and image quantification and mensuration equipment. (The photo materials were developed in a home darkroom). Before describing the objectives and results of the first study, a word is in order on how the photo materials used in the study were generated.

Three scenes, representative of the kinds of photographs image interpreters are required to interpret, were selected from Air Force reconnaissance files. One scene was of a military airfield in the Far East, another of an industrial harbor installation in Korea, and a third of a harbor in the Far East. In each case, the original negatives were of very good quality; and the scenes were judged to contain enough objects of military significanc e to insure adequate response variance.

Duplicate negatives of the three scenes were prepared on Eastman Kodak Fine Grain Safety film. The duplicates had essentially the same measured absolute density and density range as the originals. In each case, a 5 x 5-inch section of a 9 x 9-inch duplicate negative was printed by projection on Haloid-Xerox Variloid glossy photo paper, the type most commonly used currently by the Air Force Aeronautical Chart and Information Center in Washington, D.C.

The image variables introduced were photo scale, sharpness, and contrast. As indicated in Table 1, a set of 24 prints was generated from each negative by varying scale (4 levels), sharpness (2 levels), and contrast (3 levels). Variations in scale were achieved by enlargement or reduction. The standard print format sizes for the four scales of each scene were 9" x 9", 5" x 5", 3½" x 3½", and 2³⁄₁₆" x 2³⁄₁₆". The three levels of contrast were obtained by the use of contrast control filters...

Of the two levels of sharpness (termed unsharp and sharp in Table 1), the sharp level was printed by focusing the image as critically as possible using a magnifying viewer. The unsharp level was arrived at by first producing a critically sharp image as indicated above, and then introducing a diffusion filter consisting of two thicknesses of a nylon stocking sandwiched between two pieces of 2¼" x 2¼" slide cover glass. Using this projection system, the difference in sharpness between sharp and unsharp levels was checked using the National Bureau of Standards high contrast and low contrast resolution charts. Both test charts were printed on the paper at "normal" contrast with 34 lines per millimeter for the sharp level and 17 lines per millimeter for the unsharp level. These numbers in no way indicate the resolution of the original negatives; they only give an impression of the differences in sharpness introduced by the diffusion filter.

TABLE 1

CHARACTERISTICS OF STIMULI USED IN RATINGS OF USEFULNESS FOR EXTRACTING INTELLIGENCE INFORMATION

Scene A: Military Airfield, Far East 1953
Scene B: Industrial-harbor Installation, Korea 1950
Scene C: Harbor, Far East 1953

	Photo Scale			Sharpness	Contrast
Stimulus	Scene A	Scene B	Scene C	1 = unsharp 2 = sharp	1 = low 3 = high
1	1:30,800	1:16,600	1:32,100	1	1
2	"	"	"	1	2
3	"	"	"	1	3
4	"	"	"	2	1
5	"	"	"	2	2
6	"	"	"	2	3
7	1:18,900	1:10,200	1:19,700	1	1
8	"	"	"	1	2
9	"	"	"	1	3
10	"	"	"	2	1
11	"	"	"	2	2
12	"	"	"	2	3
13	1:13,800	1:7, 400	1:14,200	1	1
14	"	"	"	1	2
15	"	"	"	1	3
16	"	"	"	2	1
17	"	"	"	2	2
18	"	"	"	2	3
19	1:7,600	1:4,100	1:7,900	1	1
20	"	"	"	1	2
21	"	"	"	1	3
22	"	"	"	2	1
23	"	"	"	2	2
24	"	"	"	2	3

The contrast, sharpness, and scale differences introduced into the sets of prints were, in the judgment of an experienced image interpreter, consistent with the range of values found in operational reconnaissance photography. The same degree of magnification and reduction was applied to each scene. Further, sharpness and contrast combinations were controlled to insure that the quality of each set of prints was, insofar as possible, equivalent.

The first study explored the interrelationships among judgments of overall quality, interpreter performance, and physical photo variables. The reliability of average quality judgments and the relationships of these average judgments across different sets of photos were also explored. The 24 prints of each scene were rank ordered in terms of their overall quality by 18 image interpreters. Quality was defined for the interpreters in terms of "usefulness for the extraction of intelligence information." The average rank assigned each print by the interpreters was used as the measure of judged quality. It should be borne in mind that the prints constitute the sample. The purpose of having more than one interpreter judge the prints or identify objects on the prints was solely to increase the reliability of the quality and performance values assigned to the prints. The range of these variables depends upon the particular sample of prints selected out of the universe of possible prints.

As described above, the 24 prints made from one negative had been made similar to the 24 prints made from the other two negatives. By assuming that there was only one set of 24 stimuli which had been ranked three times (once for each scene), the correlation coefficient between averaged judged ranks could be obtained. Under this assumption, each stimulus had three scores. For example, the three scores for Stimulus 4 in Table 1 are the average ranks given Print 4 by the 18 image interpreters in judging each of the three sets of prints. Table 2 gives the intercorrelations across the 24 stimuli of the average ranks assigned to similarly generated prints of the three scenes. These intercorrelations averaged .92. Apparently, prints having relatively the same sharpness, contrast, and photo scale tend to receive the same overall rank across different content areas.

TABLE 2

INTERCORRELATIONS OF AVERAGE JUDGED RANKS OF 24 PRINTS OF THREE SCENES

(Averages based upon judgments of 18 interpreters)

	Scene A	Scene B
Scene B	.93	--
Scene C	.93	.91

Table 3 shows the split-half reliability coefficients of the average judged ranks for the three scenes. The coefficients were obtained by applying the Spearman-Brown formula to the correlation coefficients of the average judgments of 9 interpreters with those of 9 other interpreters. The high reliability indicates that very stable judgmental consensus can be obtained by pooling the rankings of several judges.

TABLE 3

SPLIT-HALF RELIABILITY COEFFICIENTS OF AVERAGE JUDGED RANKS

(Applying Spearman-Brown formula to correlation coefficients between average judgments of 9 interpreters)

	Scene A	Scene B	Scene C
r	.98	.98	.97

The correlation between the average judged ranks assigned to the 24 prints of each scene and the scale, sharpness, and contrast values assigned to the prints is shown in Table 4. Sharpness was most closely related to average judged rank. Scale showed moderate relationship, while the contrast variations introduced into the prints bore little relationship to the judged quality of the prints. The multiple correlation coefficients shown are probably inflated, owing to the artificial zero intercorrelations produced among the physical variables by our crude assignment of sharpness, scale and contrast values to the prints. For example, among the prints assigned a given scale value, there were as many prints with low sharpness as there were prints with high sharpness. The correlation between scale and sharpness across the 24 prints would therefore be zero. As Dr. Paul Roetling of Cornell Aeronautical Laboratories has indicated, image physical variables in operational imagery may have substantial correlation. Although the multiple correlation coefficients obtained between the judgements and the physical variables may tend to be overestimations, the degree of relationship found between the overall judgments of quality and the physical variables is quite promising.

TABLE 4

CORRELATION OF PHYSICAL QUALITY VARIABLES WITH AVERAGE JUDGED RANKS[a]

	Scene A	Scene B	Scene C	Total
Scale	.39	.49	.42	.43
Sharpness	.81	.77	.80	.79
Contrast	-.01	.00	.12	.04
Multiple R	.90	.91	.91	.90

[a]High rank indicates high quality.

Summarizing Tables 2, 3, and 4, we see that overall rankings of the usefulness or quality of prints can be made reliably; that prints similarly treated but from different negatives will tend to be ranked equivalently; and that such overall rankings are highly predictable from the image variables. But what of the validity of the judgments in terms of their relationship to the performance of interpreters examining the imagery?

Each of the 72 prints in the study was examined by two interpreters assigned randomly to the prints. The interpreters participating in the study had just completed a basic military officer image interpretation course. Each interpreter examined one print of each scene. The interpreters were asked to free search the three prints in order to detect and identify specified objects on the prints. Each interpreter looked at each print three times, each time searching for different objects. The objects searched for, were for the most part objects that did not require special skill or knowledge to identify. The total number of right and wrong identifications and the combined accuracy scores[1] achieved by the two

[1]Combined accuracy score = total right achieved by two interpreters divided by total right and wrong achieved by two interpreters.

interpreters examining a given print were used as the performance scores associated with the print. A set of control variables was introduced to compensate for differences in average ability levels of the interpreters examining the various prints. The average right, wrong, and accuracy scores achieved by the pairs of interpreters on three other tactical performance measures administered earlier were used as control variables.

Table 5 shows the partial correlation coefficients (with the ability of the interpreters partialled out) of the quality variables, (scale, sharpness, contrast, and judged rank) with the number right standard scores. Standard scores were used here to equate the mean and standard deviations of the number right score across Scenes A, B, and C. The correlation coefficients in the Total column in Table 5 were calculated by considering the 72 prints as one sample. Of the variables studied, scale was most highly correlated with number of correct identifications. Judged rank was generally the next most valid variable. Contrast had practically no relationship with performance.

TABLE 5

PARTIAL CORRELATION OF IMAGE QUALITY VARIABLES WITH NUMBER *RIGHT* STANDARD SCORES

(with Rights control variable partialled out)

	Scene A	Scene B	Scene C	Total
Scale	.41	.81	.53	.67
Sharpness	.38	.14	.25	.27
Contrast	.00	−.02	−.13	−.03
Judged Rank	.59	.54	.39	.48
Multiple R	.62	.83	.61	.63

It should be pointed out with regard to these correlation coefficients, that their magnitude depends to a large extent upon the range and distribution of physical values introduced into the imagery. Contrast, for example, could have been made into the dominant predictor of performance if some photos with practically zero contrast had been introduced. The problem is very similar to the restriction in range problem encountered in psychometrics – obtaining a realistic distribution of physical image values in a sample of prints is very much like defining a standard of normative population. Standard distributions of physical image variables should be established in order to compare results across different experiments and to compare magnitudes of correlation among variables within an experiment. Perhaps regression line slopes and errors of estimate should be used rather than correlation coefficients in comparing results.

The partial correlation coefficients of the image quality variables with the number wrong standard scores assigned to the prints are shown in Table 6. The multiple correlation coefficients for predicting wrong scores from the four quality variables were not so high as for predicting right scores. This is not too surprising. In our research wrong scores have, in general, proved harder to predict than right scores. Wrong scores from different performance measures have had very low intercorrelations; the correlation of interpreters' wrong scores with their aptitude scores has been relatively low. In the present study, however, the size of the multiple correlation obtained between the image quality variables and wrong scores are quite promising. The results indicate that the number of wrong identifications made in examining imagery is definitely related to the quality of the imagery.

TABLE 6

PARTIAL CORRELATIONS OF IMAGE QUALITY VARIABLES WITH NUMBER *WRONG* STANDARD SCORES
(with Wrongs control variable partialled out)

	Scene A	Scene B	Scene C	Total
Scale	.26	.16	.05	.15
Sharpness	.05	-.27	.19	-.01
Contrast	.13	.10	-.03	.07
Judged Rank	.00	-.30	.00	-.11
Multiple R	.45	.53	.46	.43

Table 7 shows the correlation between the quality variables and the combined accuracy scores assigned to the prints. In general, judged rank was more highly correlated with accuracy than were the three physical variables. The multiple correlation of the quality variables with average accuracy when taken across the three sets of prints did not hold up as well as with right and wrong standard scores. The problem of how performance scores can be best equated across performance measures involving different scenes requires careful consideration. Scores consisting of percent completeness indices, that is, number right over total number of scorable objects, were tried out as a means of equating performance. However, completeness scores were not as well predicted as were right standard scores when the scores from the three scenes were pooled. The unit of measurement for specifying image quality levels must be determined. Should some kind of performance score be used or should the index of quality be expressed in some toher terms?

At any rate, the relationships between the performance scores and overall judgments of image quality found in this study are promising. If these judgments are made systematically through the use of carefully constructed psychphysical scales, a handy means for assessing image quality may be made available.

TABLE 7

PARTIAL CORRELATIONS OF IMAGE QUALITY VARIABLES WITH *ACCURACY* SCORES
(with Accuracy control variable partialled out)

	Scene A	Scene B	Scene C	Total
Scale	.00	.34	.20	.10
Sharpness	.13	.30	-.10	.06
Contrast	-.13	-.14	-.04	-.06
Judged Rank	.25	.61	.14	.23
Multiple R	.40	.71	.49	.36

The purpose of the second study was to explore the dimensionality of judgments of quality of photos. As the quality of photos was considered almost certainly to be a multi-dimensional attribute, a multi-dimensional scaling technique was used in the study. In the technique employed—the method of successive intervals (see Torgerson 1960, p. 261)—subjects are asked to judge the similarity of all possible pairs of stimuli. The judgments of similarity were made in terms of the usefulness of the photos for the extraction of intelligence information.

Twelve photos were used as stimuli. This made for 66 pairs of photos which were mounted on large cards (10" x 20"). Twenty-one image interpreters served as subjects. They were instructed to sort the 66 cards into 7 piles depending upon the similarity of the photo pairs on the cards; that is, they were told to place in Pile 1 those pairs of photos that were most similar to each other in terms of usefulness; in Pile 7, those pairs that were least similar; and in Piles 2 - 6, pairs of photos of intermediate similarity.

The 12 photos used in the experiment were prints of the same three scenes used in the first study. The same range of photo scales (3 of original 4 levels), sharpness (2 levels), and contrast (3 levels) were used as in the preceding study. In this case, 11 copies of each of the 12 stimuli were printed for mounting on the 66 cards. The scene and quality combinations employed are indicated in Table 8.

TABLE 8

CHARACTERISTICS OF STIMULI USED IN MULTI-DIMENSIONAL STUDY OF JUDGMENTS OF SIMILARITY

Scene A: Military Airfield, Far East 1953
Scene B: Industrial-harbor Installation, Korea 1950
Scene C: Harbor, Far East 1953

Stimulus	Scene	Scale	Sharpness (1 = unsharp; 2 = sharp)	Contrast (1 = low; 3 = high)
1	A	1:30,800	2	1
2	A	1:13,800	1	2
3	A	1:7,600	2	3
4	A	1:30,800	1	3
5	B	1:7,400	2	2
6	B	1:4,100	1	3
7	B	1:16,600	2	1
8	B	1:7,400	1	2
9	C	1:7,900	1	1
10	C	1:32,100	2	1
11	C	1:14,200	2	2
12	C	1:7,900	1	3

The judgments of the interpreters were punched on IBM cards and comparative distance, absolute distance, scalar product, and factor matrices for the multi-dimensional analysis were calculated on a 7090 computer using a program developed by APRO. The minimum dimensionality of the set of photos was found to be six—additive constants calculated for fewer dimensions yielding imaginary roots for the comparative distance matrices or negative values in the absolute distance matrices.

The unrotated factor matrix obtained from the analysis is given in Table 9. The first two factors were readily interpretable from their relationships with the physical variables. Stimuli with positive loadings on Factor I had relatively large photo scales; stimuli with negative loadings on Factor I had relatively small photo scales. The average loadings on Factor II of stimuli with high sharpness is +.76; the average loading on Factor II of stimuli with low sharpness is -.76. These two factors had by far the largest latent roots. Apparently, the photo scale and sharpness of the stimuli strongly influenced the judgments of quality. This finding was corroborated by the high correlation found in the first study between these same physical variables and the judgments of overall quality.

TABLE 9

UNROTATED FACTOR LOADINGS OR SCALE VALUES OBTAINED FROM PRINCIPAL AXES ANALYSIS OF SCALAR PRODUCT MATRIX OF ABSOLUTE DISTANCES

					Dimensions or Factors					
Stim.	Scene	Scale	Sharp.	Contr.	I	II	III	IV	V	VI
1	A	1	2	1	-2.21	-.36	-.12	.04	-.96	.19
2	A	3	1	2	.99	-1.23	.20	-.45	-1.17	-.60
3	A	4	2	3	1.31	1.04	-1.87	.73	.04	.22
4	A	1	1	3	-.97	-1.81	-1.53	-.51	-.77	-.45
5	B	3	2	2	-.17	2.22	.05	-.37	-.04	-.34
6	B	4	1	3	2.10	-.50	.27	-.31	-.18	.27
7	B	1	2	1	-1.97	-.21	.56	.35	.38	.10
8	B	3	1	2	.58	-.33	.78	-1.39	.73	.40
9	C	4	1	1	1.39	-.12	1.07	1.24	.57	-.71
10	C	1	2	1	-2.35	-.32	.44	.75	.06	.31
11	C	3	2	2	-.49	2.17	.05	-.59	-.04	-.13
12	C	4	1	3	1.78	-.55	.10	.52	-.14	.73
				Latent Root	28.04	16.46	8.24	5.99	3.95	2.15

The remaining factors proved more difficult to interpret. In order to determine whether factors associated with the contrast or content of the stimuli could be found in the data, an orthogonal rotation was performed on the remaining factors. In the rotation, a least square fit was sought with hypothetical contrast and content loadings. (The hypothetical contrast loadings used were 1, 2, and 3 as in Table 9. The stimuli involving Scene A were given hypothetical loadings of −1, and the stimuli involving Scenes B and C were given loadings of +1 on the hypothetical content factor). The results are shown in Table 10. Fair success was achieved in demonstrating the existence of contrast and content factors. Stimuli involving Scene A have negative loadings on Factor III while generally positive loadings on Factor III were obtained for the stimuli involving Scenes B and C, the harbor installations. Similarly, the loadings of high contrast stimuli were positive on Factor IV, the loadings of low contrast stimuli were negative. Perhaps, if the judgments of more interpreters had been available, these dimensions would have been more clearly defined.

TABLE 10

ROTATED FACTOR LOADINGS OR SCALE VALUES OBTAINED FROM PRINCIPAL AXES ANALYSIS OF SCALAR PRODUCT MATRIX OF ABSOLUTE DISTANCES

	Stimulus Characteristics				Dimensions or Factors			
Stimulus	Scene	Scale	Sharp.	Contr.	I	II	III	IV
1	A	1	2	1	-2.21	-.36	-.51	-.05
2	A	3	1	2	.99	-1.23	-.82	-.45
3	A	4	2	3	1.31	1.04	-1.07	1.03
4	A	1	1	3	-.97	-1.81	-.83	1.07
5	B	3	2	2	-.17	2.22	-.17	-.10
6	B	4	1	3	2.10	-.50	.20	.07
7	B	1	2	1	-1.97	-.21	.65	-.44
8	B	3	1	2	.58	-.33	1.04	.48
9	C	4	1	1	1.39	-.12	.74	-1.44
10	C	1	2	1	-2.35	-.32	.51	-.37
11	C	3	2	2	-.49	2.17	-.08	.11
12	C	4	1	3	1.78	-.55	.37	.13

The results of these two studies point, we believe, to the promising utility of judgments in assessing image quality. Judgments of quality apparently reflect the physical make-up of the imagery. The average judgments of a number of interpreters are certainly reliable and valid. Future research into the psychophysical aspects of image quality can take many possible directions. It was not the intent of this presentation to indicate a specific research approach. The presentation was intended, rather, to point out the potential fruitfulness of research in this area.

REFERENCE

Torgerson, Warren S., *Theory and Methods of Scaling*, John Wiley and Sons, Inc., New York, 1958.

GENERAL DISCUSSION

Dr. Roetling (CAL):

If I understand you correctly, on your paired comparison, you did get cases in which the interpreters rated pictures as very much alike despite the fact that you had gross differences in the physical variables.

Dr. Sadacca (APRO):

You must remember what the judgment of similarity was. It was similarity in regard to usefulness in the extraction of intelligence information, not similarity to the judged physical characteristics.

Dr. Roetling (CAL):

In that case, you have demonstrated the point that Mr. Cook brought up earlier; namely, that with gross differences in variables, you can still find cases where the interpreter can be trained in a simple test to rate these as comparable in terms of performance or interpretability. In fact, we may get the same value for summary measure. Essentially what you are rating here is interpretability for grossly different looking pictures. This means a compensation of physical variables that can give the same answer from two sets of variables.

Mr. Speer (Houston-Fearless):

I wonder if you would explain what your definition of contrast is? Is this a high contrast print versus a low contrast print or a contrast of particular objects?

Dr. Schwartz (APRO):

It was a high-low contrast comparison of the points within the prints themselves. I was looking at white and very dark areas but there was no means of measuring these observations. On the other hand, people consistently arrive at similar judgments irrespectively of how I arrived at it. It did not seem to be a very important factor as we have expressed here in the performance of the subjects.

Mr. Devoe (U of Mich):

I have some supporting evidence on the same area, mostly in reading MGI radar films. We made some measurements of signal-to-noise values on the film. It is quite obvious that there is a big difference in the film itself. Although some operators do not like some films they do not seem to be in any great disagreement about their dislikes. In any case you do get the same results. It does not affect their performance appreciably. This same principle seems to carry through on many different forms of display presentations.

A STUDY OF IMAGE QUALITIES AND SPEEDED INTRINSIC TARGET RECOGNITION

Dr. Samuel H. Winterstein
International Business Machines Corporation
Space Guidance Center
Owego, New York

ABSTRACT

A brief summary is presented of an experimental Image Quality Independent Research and Development program designed to determine the relationships between image quality and military target recognition performance. An explanation is given of the methodology employed and the investigations conducted during the program. The major results are described and the present status of the effort is indicated.

In this program, which is company-funded and government-funded under an IRAD program*, we were interested in investigating whether one could predict recognition performance from measurable image qualities – from physical measurement of the images. There are two approaches to this problem: (1) look at simple abstract figures, and manipulate these figures experimentally to learn the basic phenomena, or (2) work with the complex displays, trying to discover the determinants under realistic conditions. The latter approach was the one taken because this approach could yield some results sooner.

To accomplish this end, two areas were studied:

1. An image quality measurement program was established to measure physical image qualities.
2. A recognition study was initiated and carried out to find the probability of recognition of various targets under various scales and photograph degradations.

Briefly the program was as follows:

Aerial photographs were selected and degraded, and image quality measures were defined and applied to these photographs. The recognition study obtained recognition probabilities on these degraded photographs, and these were compared to the predictability of recognition as determined from the measured image qualities.

In the recognition study, eight targets were used: airfields, nuclear storage sites, barges, fighters, industries, towers, antennas, and trucks. These were studied at three scales: 1:2000, 1:10,000, and 1:50,000 (Figure 1). Images were degraded in contrast, resolution, and grain (Figures 2-4). Resolution degradation of the images was accomplished by a process of off-contact photographic printing. The grain degradation was accomplished by photographing a mask of sandpaper and combining a sandpaper mask photographically with the images. (Figures 1, 2, 3 and 4 are examples of scales, targets, and the levels of degradations that were used).

* IRAD – Independent Research and Development.

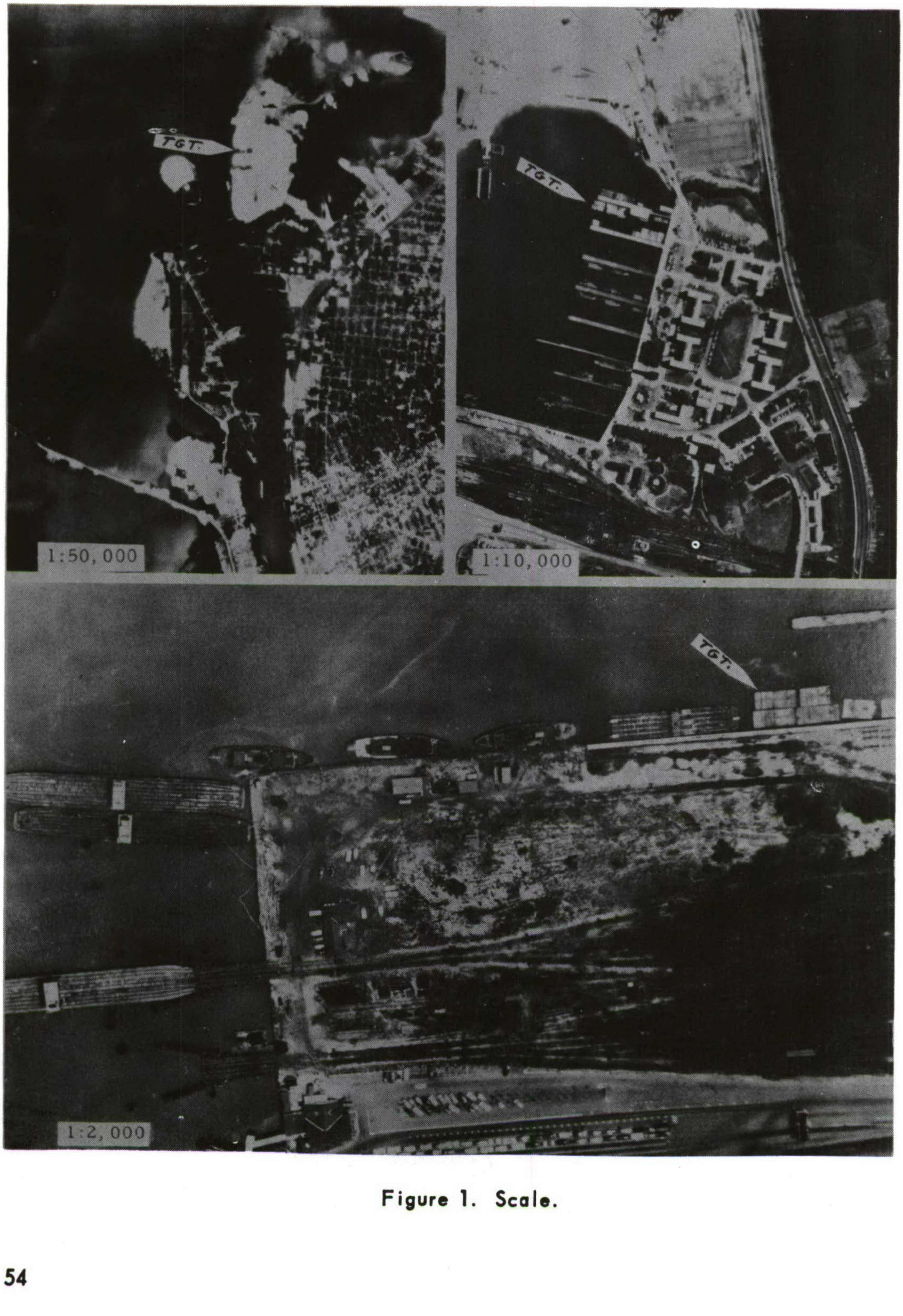


Figure 1. Scale.

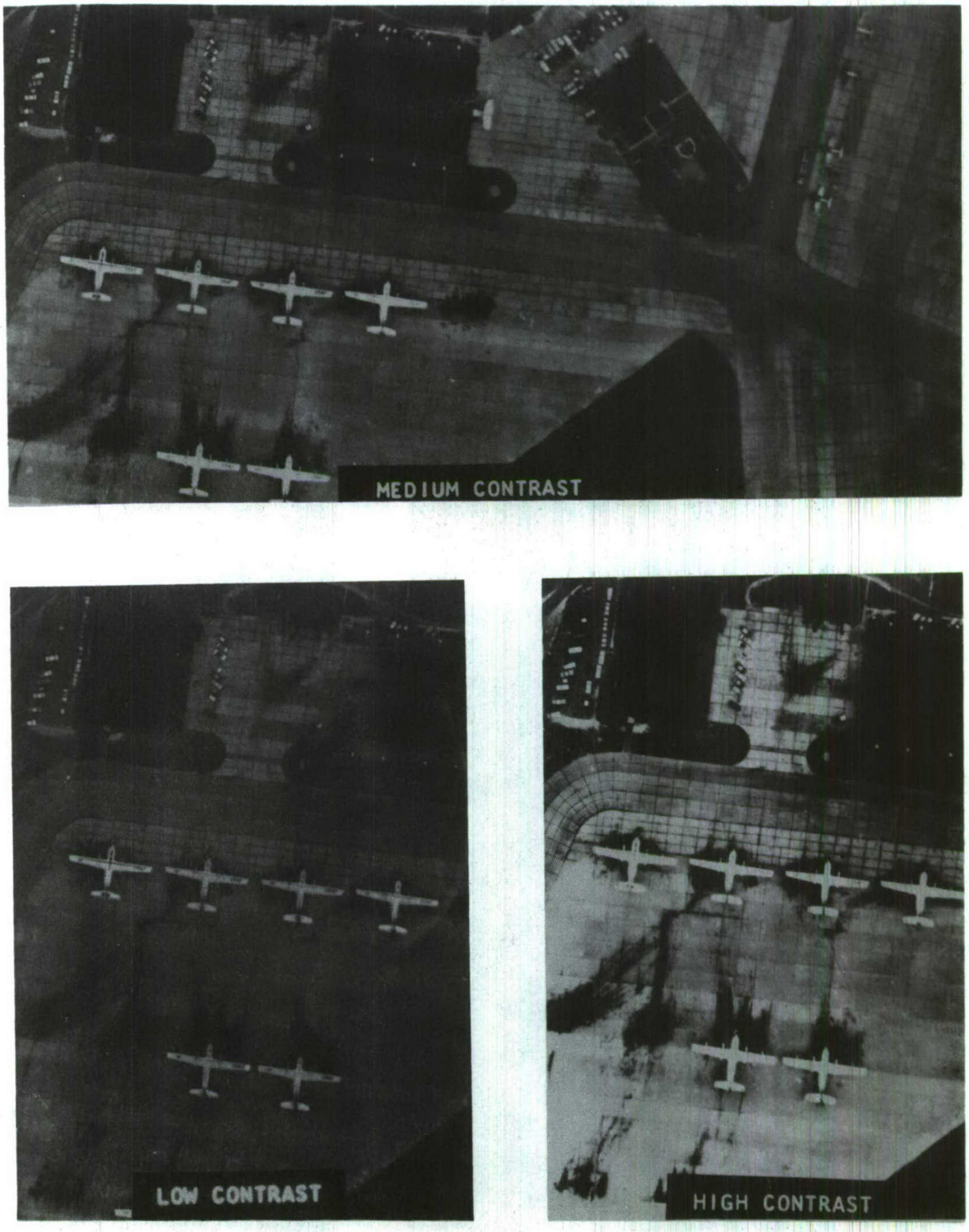


Figure 2. Contrast.

Figure 3. Resolution.

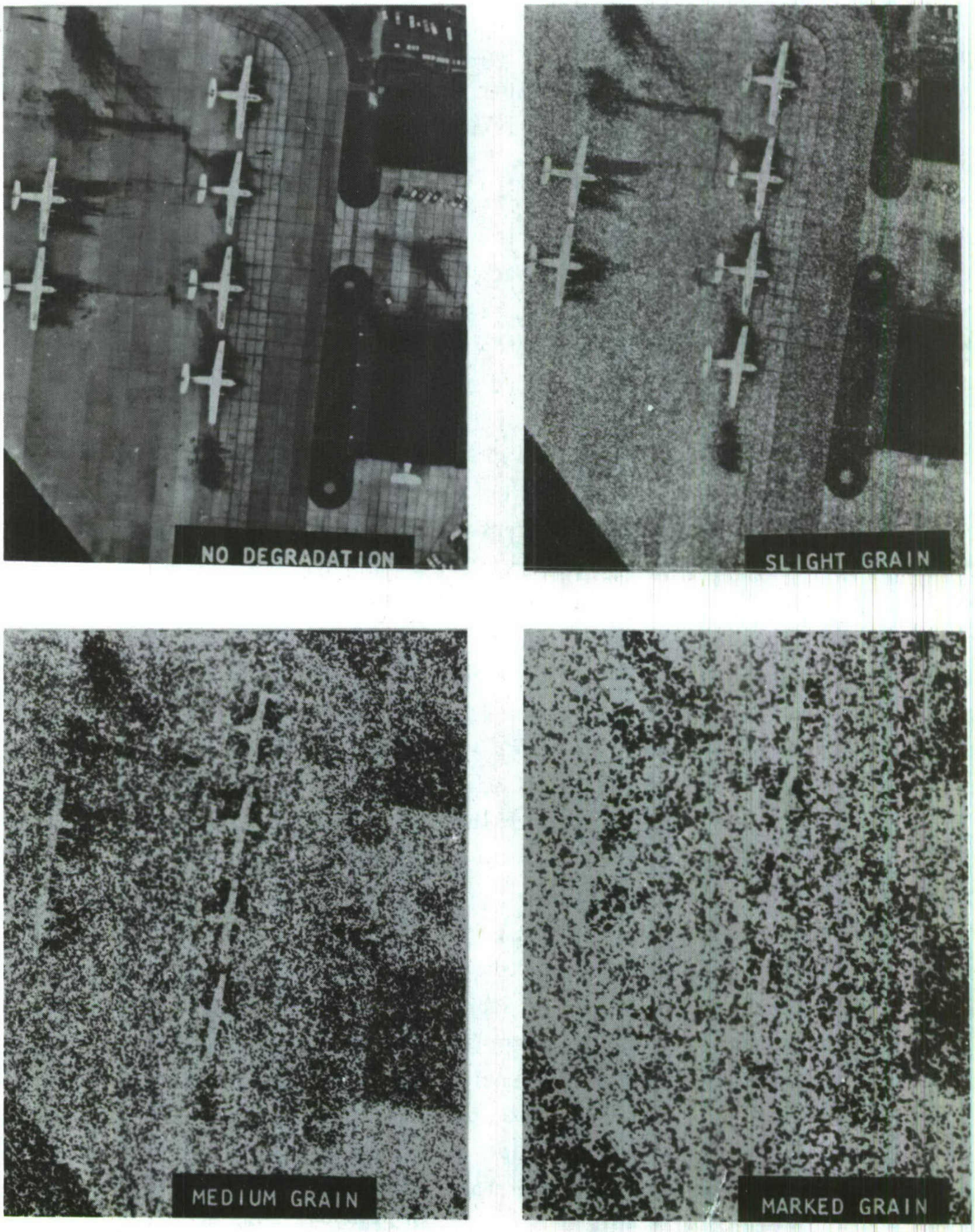


Figure 4. Grain.

In the experimental design, sixteen different combinations of the variables were selected, including the extremes of contrast, grain, and off-contact (fuzziness), with several of their combinations. The equipment used for this testing was an analog-to-digital converter. This equipment provided a light box with cross hairs where a subject could place the cross hairs on a photographic image to approximately 0.010 inch accuracy for target location. The equipment was connected to an IBM card punch. In operation, a subject placed the cross hairs on a target, went to the keyboard of the card punch, named the target, and recorded how confident he was that he had located the target correctly. The equipment recorded the cross hair position, the name of the target, the subject's confidence level, and the time of response. (A timer was attached to the equipment so that for every response, a time interval was recorded).

The subject had to locate one main target and other targets for each picture, i.e., he was given a photograph and was told that this was a photograph of an airfield, and possibly more than just an airfield. He was required to locate the airfield and any other targets (in the list of targets mentioned previously). Twenty-four subjects were trained and tested. All were laboratory personnel having previous experience in target recognition studies. Each subject saw each one of the 16 degraded pictures. However, he never saw the same scene under more than one degraded condition.

Results of the recognition study showed the overall probability of recognition to be about 0.26. However, this ranged from 0.88 for airfields down to about 0.06 for antennas. The contrast variable seemed to have no effect on probability of recognition. Off-contact and grain conditions produced drastic effects. The confidence measure indicated the ability of a subject to tell how accurately he could perform. All subjects appeared to be over-confident. In other words, they all said they did a little better than they actually did.

In the measurement program, density measurements were taken with microdensitometer traces of the pictures, with the output recorded on punched cards. From these basic density measures, simple image quality measures were derived for resolution, contrast, and grain. These measures were based on local maximum, local minimum, and slope characteristics of the output trace. Distributions of these characteristics were formed, rating techniques were applied, and these measures were correlated with measures of probability of recognition and with each other. Multiple correlations based on linear and quadratic terms of these quality measures with recognition performance was about 0.9 for eight predictor variables. By reducing the number of predictor variables to four, the correlation between recognition performance and the measures was 0.86. These methods are successful for the off-contact printing method of degradation. However, some difficulty has been experienced with the grain conditions. In general, results of these studies demonstrate an ability to quantitatively predict recognition performance from image qualities, although there is a need to cross-validate these results.

At present, efforts are continuing on the program in such areas as refining the scanning equipment, analyzing recognition data more fully, and scanning more pictures for evaluating some of these results.*

* Since this presentation, a final report has been written on the results of this program—"A Study of Image Qualities and Speeded Intrinsic Target Recognition: Final Report—Abstract and Summary of Conclusions," C.A. Bennett, S.H. Winterstein, J.D. Taylor, R.E. Kent. IBM No. 63-535-1. Feb 5, 1963.

GENERAL DISCUSSION

Mr. Pickering (Eastman Kodak):

This is more in the line of a comment than a question. At the SPSC meeting last Spring, one of the gentlemen was talking about the degradation of images in similar manner that you were using here. In his particular experiment, he had used a shaded patch as the light source, that is, the light source was a broad diffused radiator, and over this was placed a photographic pattern giving him a shaded patch so that he could, in effect, control the contour of the spot image, enabling him to know the degradation generated. This might be another variable that you would want to control in a similar test.

Dr. Winterstein (IBM):

The degraded conditions that we used were rather artificial, but on the other hand we feel that they looked quite realistic.

SECTION III

COMPARATIVE COVER ANALYSIS

FACTORS AFFECTING CHANGE DISCRIMINATION

Dr. Conrad Kraft
Dr. C.L. Klingberg
The Boeing Company
Seattle, Washington

ABSTRACT

A background research program at Boeing is described which is relevant to an ongoing study on photographic change discrimination. The objectives of the present investigation are to determine the effects of selected presentation variables on the detection and identification of changes in comparative-cover photo pairs. In particular, the relative advantages of a side-by-side versus an apparent motion mode of display is being investigated. A detailed description of the research design is presented and preliminary findings are described.

BACKGROUND INFORMATION

Dr. Kraft:

The recent Cuban situation would serve as a convenient introduction to this discussion of target change detection in comparative cover photography; but I would like to take you back about three years when we, as Engineering Psychologists, were attempting to demonstrate the need for this type of research to a large aircraft company, one that is known for and probably still is oriented toward the design and construction of large airframes.

The concept we advanced was that there had not been an equal distribution of effort in developing all phases of the reconnaissance system. The major technological effort had been directed toward the collection and storage aspects of the system. As a result we have the capability to collect pictures at high rates and to store and withdraw information at equally excellent rates; and certainly the needers' requirements are large and immediate. Two phases that had received much less attention were the classification and encoding of pictorial data before storage and, after retrieval, the task of image interpretation. This unequal distribution of the technological effort has made the reconnaissance system a victim of the "dumbbell effect." (Figure 1)

The adjective "dumbbell" is intended not as a disparaging word but as a descriptive word: descriptive of the flow of information through the reconnaissance system of today. The top ball of the dumbbell – collection capability – is large, as is the handle which represents the storage capability. The need for information creates the equally large lower ball of the dumbbell. The flow of information, however, can be no greater than whatever improvement we can make in the two restrictions shown as "classification" and

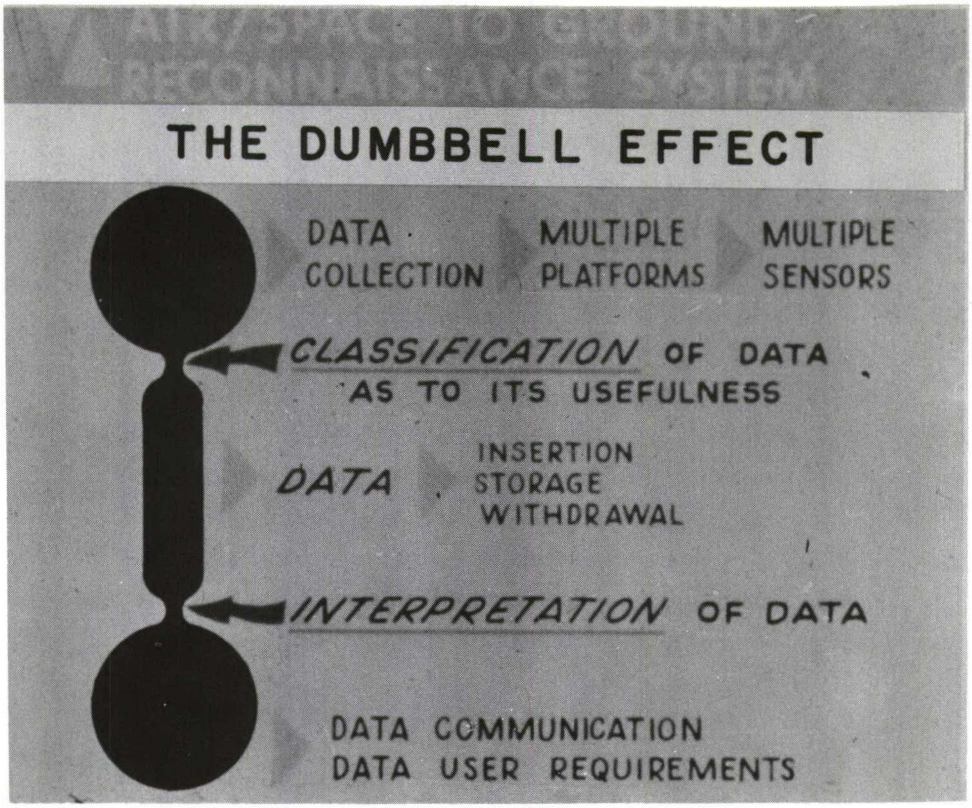


Figure 1.

"interpretation" capability. The magnitude of the restriction may be illustrated by the next slide. (Figure 2)

The first portion of the information in this slide is based on the capability of Tiros I, our first weather satellite, based on information from *Aviation Week*. Along the abscissa are the years 1950 to 1970. Tiros' capability appears in 1958; and if this satellite had lasted a year instead of 13 weeks, it would have taken 1,056,000 pictures. It is possible that by 1970 we could have 20 satellites in orbit with four sensors in each. In 1970, then, the number of pictures from space could total 56,000,000 per year. The two triangles are representative of the pictures that could be obtained from within the atmosphere. The lower triangle represents the "basic coverage" needs of a single "Field Army" of a 100-mile wide, 300-mile deep, active combat area. If such an army were engaged with a very active and mobile enemy, it would probably wish to repeat this basic coverage on a daily basis: a requirement for 144,000 pictures per day. The upper triangle represents the number of pictures required by ten Field Armies – the number that were active in Europe in World War II. A midpoint between these two triangles is equal to the number of pictures which could come from the 20 satellites. Employing Jay Enoch's data, an estimated one-half hour per picture is required for technical interpretation; one finds that each year of collection would require approximately 48,000 man-years for interpretation.

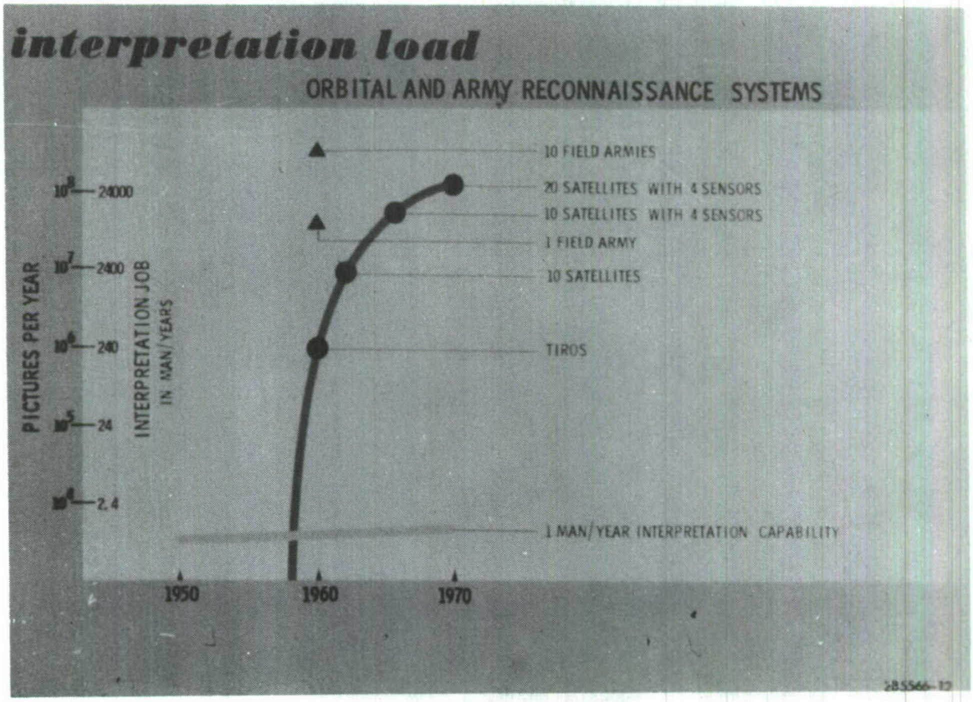


Figure 2.

Interpretation of imagery in the quantities discussed above are obviously beyond the total capacity of all the trained personnel in the United States. We feel that the solution to the problem does not lie in merely training enormous numbers of interpreters, not does it lie entirely in automation. The solution lies in better methods of utilizing men and machines, and particularly in the utilization of man's unique abilities to integrate information both spatially and temporally.

These abilities, as applied by man in the imagery-interpretation task, are demonstrated in the next slide (Figure 3). This illustrates the integration of previous knowledge with the physical aspects of the light and shade in imagery to allow interpretation of this photograph as the area of the Red Sea, the Nile River, and the Suez Canal.

The next slide (Figure 4) demonstrates man's ability to rapidly impose new strategies in the process of gaining meaning from imagery. This photograph could be a high-altitude, aerial view of the South Pacific. It is not, however; it is merely a picture of a Holstein cow taken with a child's box-camera, too close to the subject and slightly out of focus. To see it, one normally must change his way of looking and thinking about the picture. For example, the cow is broadside to the viewer, her head turned toward the viewer; the ears are black; the muzzle is in the lower left; and one eye is in the shadow, while the other is shaped like a horseshoe magnet. Based upon prior knowledge and this information, most viewers can organize this illustration into a picture of a Holstein cow. To provide this capability in a machine would require a huge and very complex piece of equipment.

Figure 3.

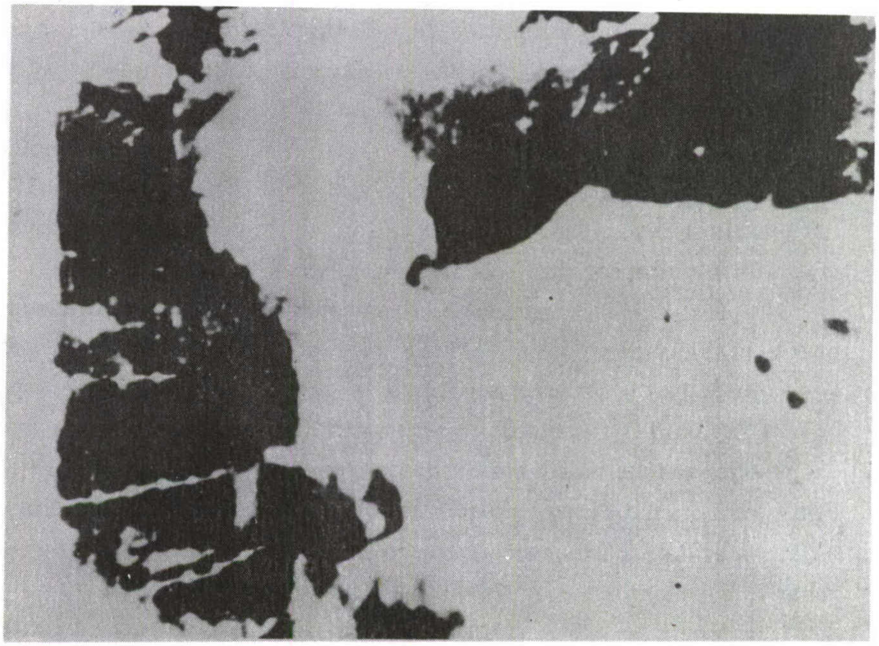


Figure 4.

Machines, however, possess capabilities beyond those of man in memory, rapid handling of digital information, rapid transformation of data, and can perform many special tasks where the human is often misled. As an example of this latter point, the nest illustration (Figure 5) presents what would appear to the individual as a series of spirals. However, a machine would not "see" it this way but rather as the series of concentric circles which it is. In other words, the optimum system would relegate to man that which men do best; and it would relegate to machines that which machines do best.

If the above postulate is applied to current methods of photo interpretation, a simple and direct approach would be to utilize semitrained personnel, operating modest but rugged equipment, to detect rapidly and accurately any changes that have occurred in comparative cover photographs (two photographs of an area taken at different times). This would allow the skilled, highly-trained photo interpreter to concentrate on the changed areas, where the highest probability of obtaining significant data exists.

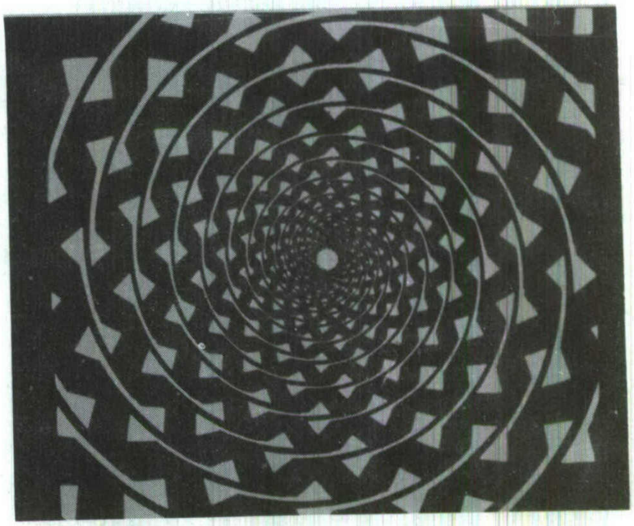


Figure 5.

We have been interested in investigating three methods for detecting changes in photographic imagery. The "side-by-side" method is the current operational technique and is included in this experimentation as a baseline condition. The "overlay" method is the first experimental one; it consists of optically mixing projected positive and negative transparencies in "slight-offset" registry wherein the illuminance is attenuated by the densities of the two photographs. The second experimental method is designated as "apparent motion," so called because any change occurring in the imagery will have apparent motion for the observer. This method presents the early and late samples in registry alternating at 1-1/2 cycles per second. To illustrate how pictorial data would appear when presented by these three methods, the following is presented. (A movie was presented at this point to illustrate how these changes appear in an air-to-ground photograph under the three techniques previously discussed.) This film sequence demonstrates the potential of the two experimental techniques. To quantify the true efficiency of these presentation methods constitutes the major experimental problem. The equipment, methods, results, and their implications follow.

The instrument shown in Figure 6 was designed principally to study the differences among the three methods of presentation as they affect ability to detect change in pictorial data. The black box in the upper right, the display system, measures 2 x 5 x 5 feet. On the stand to the left is the readout equipment which is used with this instrument. (The movie sequence was referred to again to show the functional diagram of this black box and procedure in using this equipment. The procedure is described in the next three paragraphs.)

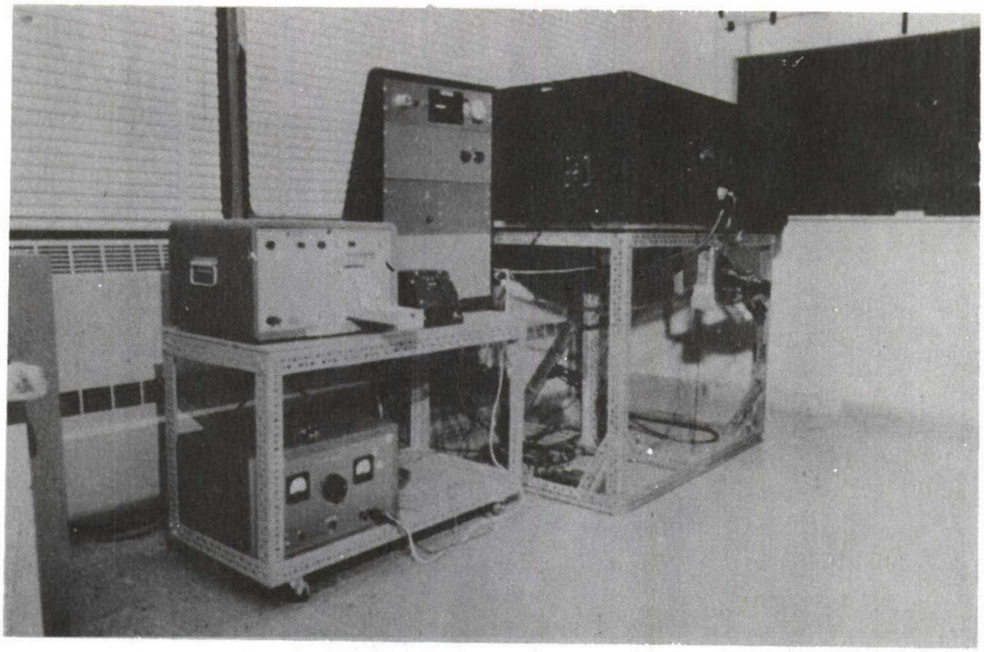


Figure 6.

This is a functional diagram, as though of a bird's-eye view of the three optical pathways. The light source in the upper right corner is contained within an aluminum sphere which is drilled to give two point sources. The light emitted by these holes is chopped by an episcotister and spread by minus lens so that it forms an area source on an opal glass. This area source is reflected by a second surface mirror, picked up by a Fresnel lens, passed through the slide, and reflected off the 50-percent mirror. A 3-1/2 inch diameter luminous cone allows a binocular viewing situation. The second pathway is a complete duplication of the first and starts from the upper right and proceeds horizontally and is turned downward by the mirror. A third channel at the lower left provides a visual feedback giving the observer information as to the correctness of reporting his selection of the changed area.

"We used two experimenters in this case. Mr. Davis is loading the instrument with slides. We cooled it by vacuum. The person in the center is the observer who is told to look in, detect the change and report it with a right-hand response. When we used synthetic stimuli, this is what the observer saw. Note the blooming effect; it represents the appearance of a change in size, the rocking is misregistration. As the observer responds she can turn on a matrix and then as she pushes the button the yellow light comes on to indicate the area that is represented by the response. If the observer has responded incorrectly, he or she has 0.5 second in which to correct the response before the automatic recording is activated.

"The second experimenter on the right checks the accuracy of the recording equipment, indicates the trial number." (End of movie sequence)

The experimental design (Figure 7) used in this first study, a 3 x 3 x 3 factorial, permitted the study of more than one variable at a time as well as the interaction among

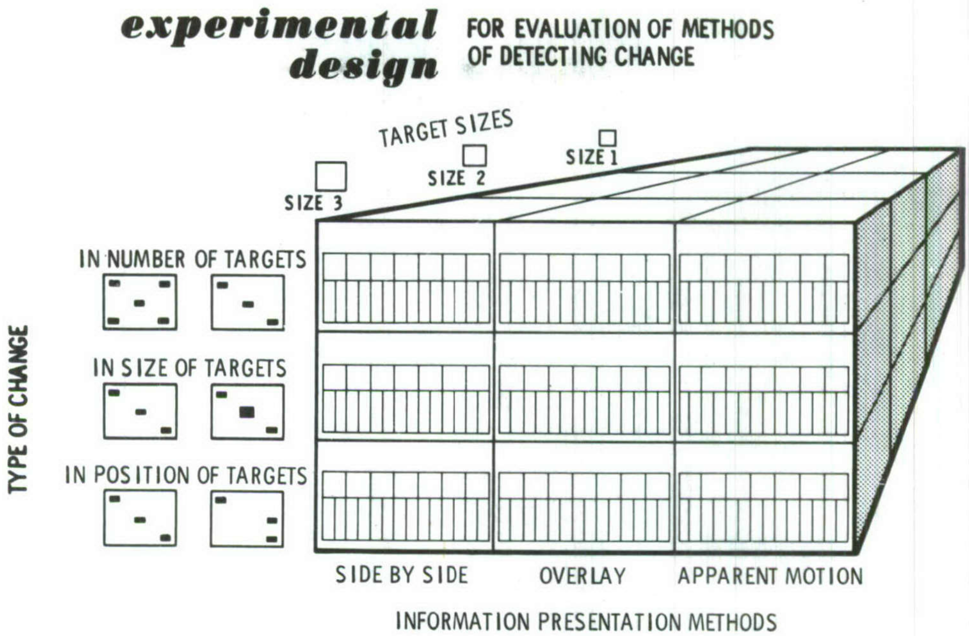


Figure 7.

variables as they affect the performance in the detection of change. The three presentation methods are shown on the abscissa, while the ordinate represents three of the five possible ways of classifying types of change. Changes in number, size, and position were also chosen as variables for investigation. Changes in configuration or color have purposely been omitted from the present study. The "Z" axis of the slide represents the object size and is related to the determination of display scale. Size 1 was four minutes of visual angle; size 2, 16 minutes; and size 3, 32 minutes. The design includes twenty-seven individual test cells.

The next slide further defines the individual makeup of one of these cells. (Figure 8) Two presentation positions were used in which the slides were placed right side up and 180° reversed. Each slide presented to the observer contained either zero, two, three, or six changes; however, the exact number was unknown to the subjects.

Fifteen observers, trained on each of the three methods of presentation, but not trained photo interpreters, were used. They reported the detected changes by pushing buttons in a 6 x 6 readout matrix. The target location, as indicated by the observer and his time of responses were recorded by a digital printer. This response system required an average of 8.5 seconds to report three changes per slide.

The dependent variables for these studies were two: (1) the amount of time in seconds to detect the average of three changes per slide, and (2) the error scores. These error scores were divided into two categories: the number of unreported targets, and the number of false reports of targets that were not changed.

(A movie sequence shown at this point demonstrates the interaction among the three methods of presentation of change and the three types of change.)

experimental design: FOR INDIVIDUAL CELL

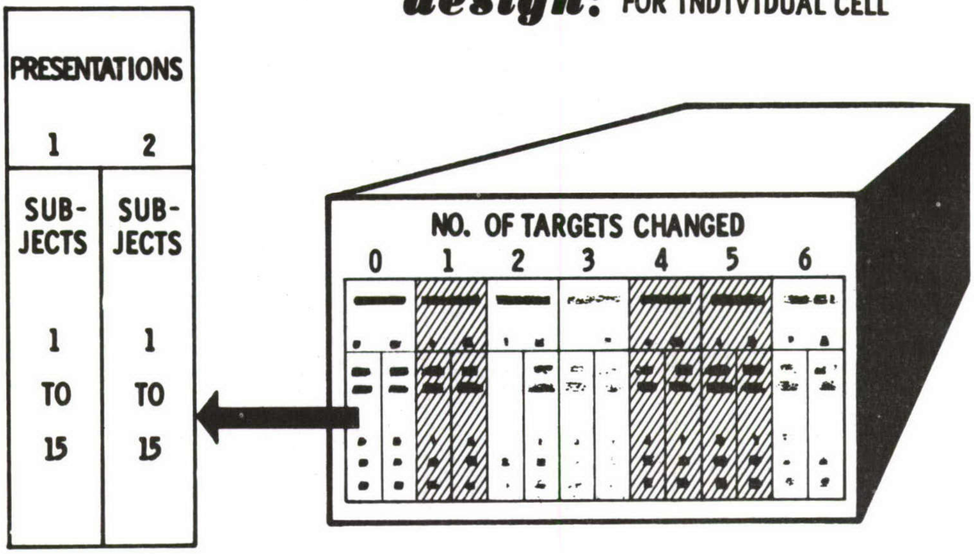


Figure 8.

Synthetic targets were used in this study and the 64 targets in each test slide presentation were randomly distributed over a matrix of 144 possible locations. There were an equal number of the three sizes of squares and there were 72 different slides. Each observer saw 72 slides under each of the three methods. For those slides involving size and position changes, the actual movement or increase in size of the target was maintained at one-half of the principal linear dimension of one side of the target. The use of these synthetic targets for study allowed precise control of the stimulus material. It was assumed that, if large differences were not found among the three presentation methods using the synthetic targets, no further studies would be undertaken. If, on the other hand, the results showed significant difference, some of the bigger problems such as registration and rectification would be studied.

The differences between the three methods proved to be highly significant and some of these results are shown in three-dimensional form. (Figure 9) For the moment, the individual heights of the nine columns may be disregarded and the average column height for each of the three major groups considered. Note that the side-by-side group is about three times as high as for the two experimental methods, with the actual means for the three groups being 58.5, 19.7 and 22.5 seconds, respectively. The interactions in the "side-by-side" method between types of change and display size were found to be not significant. Apparently, the observers frequently gave up when they assumed the changes were almost impossible to find. For the overlay and the apparent motion methods, the interactions between types of change and the display size as they influenced the total time to detect the targets were highly significant.

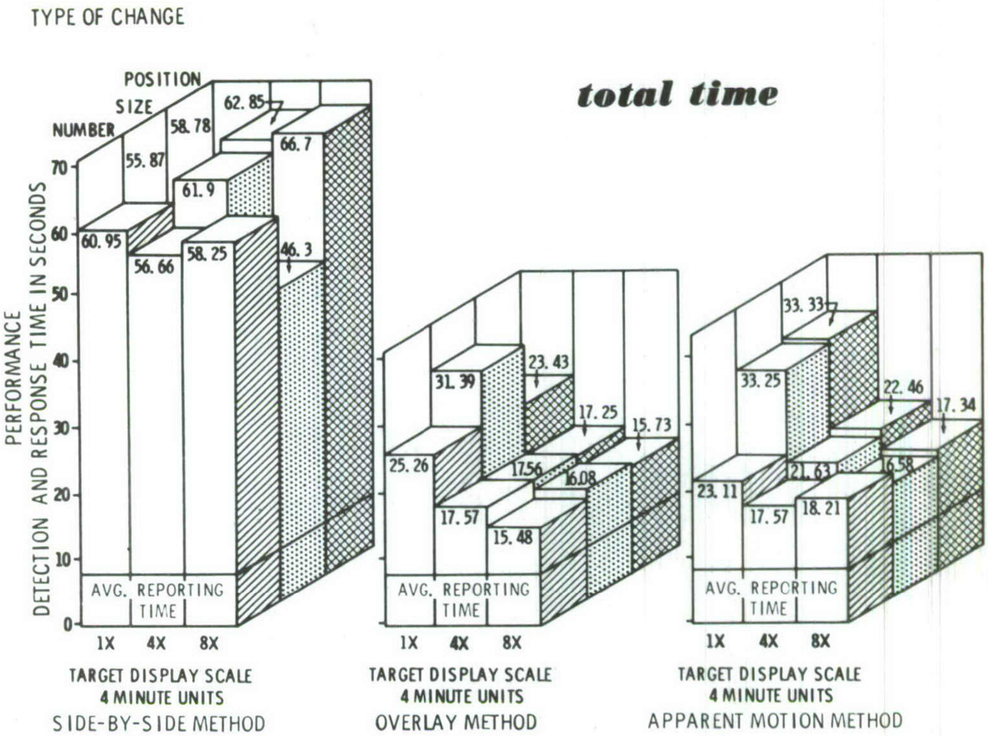


Figure 9.

This is further illustrated in Figure 10 which shows the omission errors for the various methods of presentation. In the side-by-side method and targets of 4 minutes in size, the height of the histographic representation indicates that in the most difficult type of change to detect – that of change in position – one could expect 99.1 percent errors. This error could be decreased to 60 percent by increasing the display size four times, or to 37 percent by increasing the display size to eight times. For changes in addition of targets, the distribution is almost rectangular and variations in display size do not produce a significant difference in errors of omission, Considering the overlay and the apparent motion methods, one notes that 16 minutes of visual angle in display size is optimum.

You may rightfully ask the next question, "How will these techniques work with aerial photography?" This question was at least partially answered in our second and third experiments. The next slide schematically represents the experimental designs for the second and third experiments. (Figure 11) The only basic difference is the inclusion of a condition of low visual noise and high visual noise aerial photography. The method of generating these stimuli is shown in the next movie sequence. These are actually aerial photographs kindly supplied to us by the CIA. There were 136 individual photographs of which we used 36. The low visual noise photographs were those which represented pictures taken from the same point in space. (A movie sequence was used to illustrate.) The identical reference point in space was accomplished by making enlargements 2.4 times that of the 9 x 9-inch negatives and making two prints, introducing changes to only one and then rephotographing both on the same scale.

omission errors

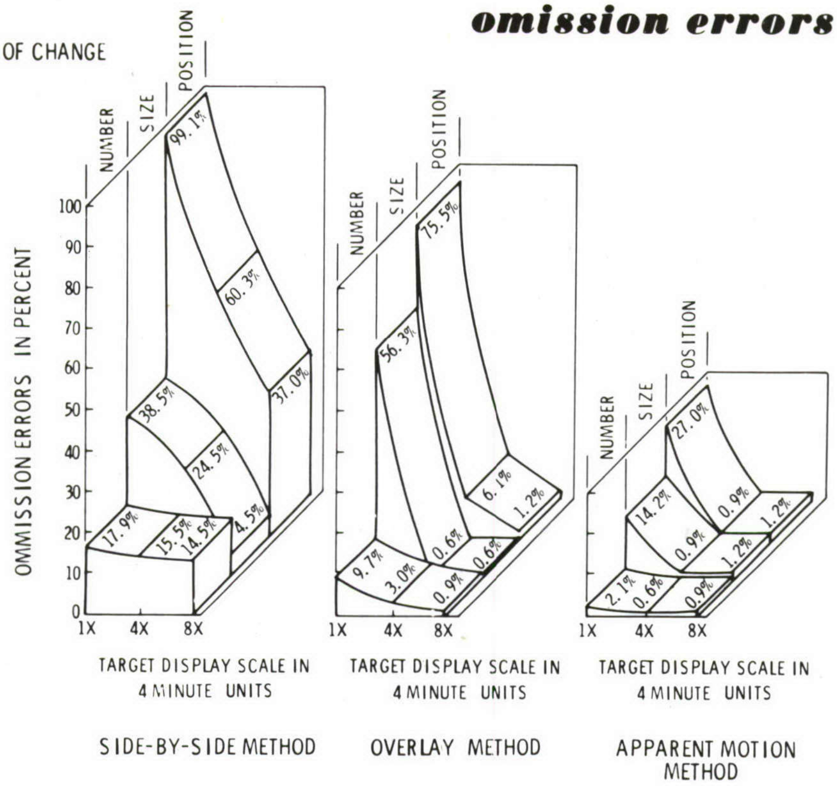


Figure 10.

experimental design II aerial photographs

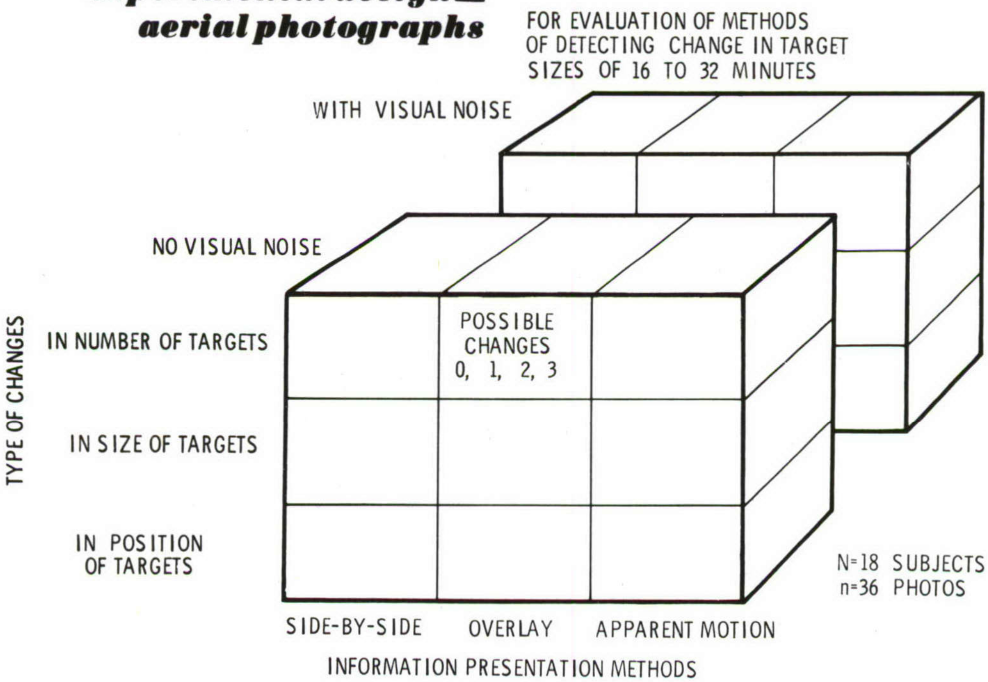


Figure 11.

The high visual noise photography is as though you did not have the ability to go over exactly the same point, but missed it in some degree – either in x or y. The overlapped area, of course, is what was used in the second study. This actually averaged about 58 degrees difference as measured from the ground reference point. This technique produced the correlated noise characteristics of the operational problem. (A movie sequence was shown here. The movie sequence illustrated photographs of low and high noise as seen with the side-by-side and apparent motion display methods.)

This slide (Figure 12) represents the results of all three studies with the dependent measure of time on the left and errors on the right. The front row represents the data obtained with the synthetic material, the middle row is data from the limited-noise photographs, and the back row is data from the high-noise pictures. As far as detection time is concerned, the general shape of the function as a product of display method was replicated but the absolute times were not. The aerial photographs took approximately 154 seconds with the side-by-side method, as compared with 60 seconds for simple forms. This is the average time per observer to detect 1-1/2 changes in a 4 x 4-inch aerial photograph at a scale of 1:15,000. The overlay method under the same conditions took about 60 seconds and the apparent motion, about 46 seconds. Accuracy performance did not generalize from the synthetic stimuli to aerial photographs. The overlay method resulted in more errors than the side-by-side method. The explanation for the poor performance is that the phenomenon of "Titchener's scintillation effect" and the "bas-relief effect" was not apparent with the complex aerial photographs. It was at this point in our research that we dropped this method from further investigation.

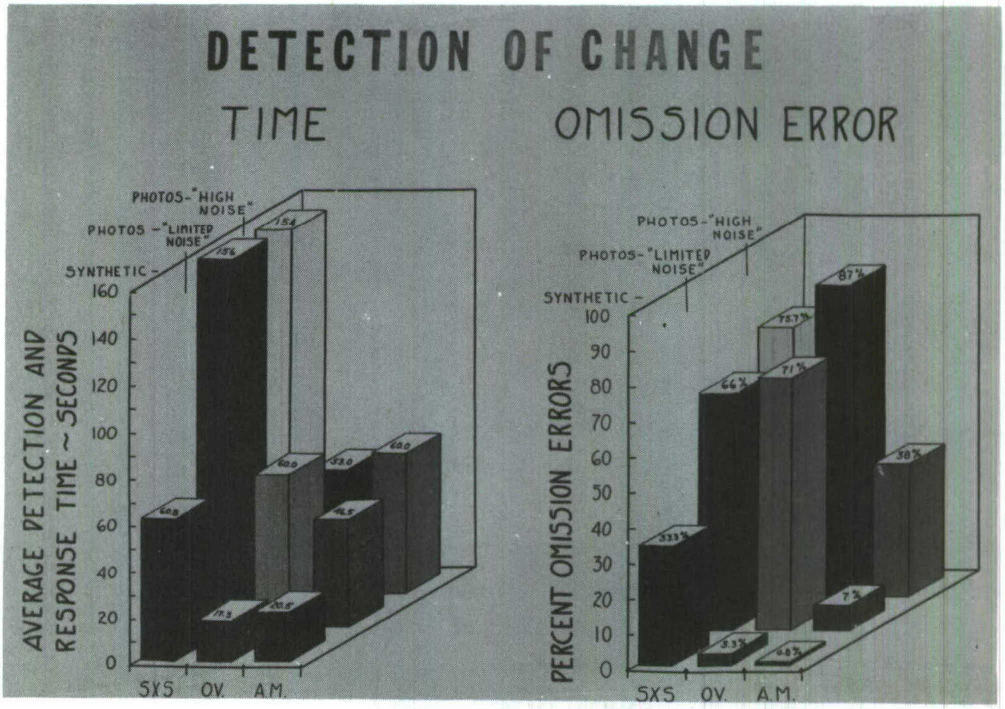


Figure 12.

The apparent motion presentation method still represented a distinct advantage over the side-by-side method for the observer in detecting change in comparative aerial photographs. The difference in error scores for the identical photographs and the systematic changes is 7 percent as opposed to 66 percent for the side-by-side method. In the case of the high-noise pictured, the side-by-side display method produced almost 76-percent error. Performance with the apparent motion method yielded only 38-percent error. The apparent motion method maintains a 2 to 1 advantage under the severest condition tested to date.

The point that Mr. Zimmerman alluded to yesterday is that current operational photography and the use of the side-by-side method should not lead to more than 25-percent accuracy in detecting changes.

In Figure 13 are the results as they pertain to commission errors. Within our experimental situation, none of the display methods produced false reports per picture greater than the 1.2 indicated for the overlay method.

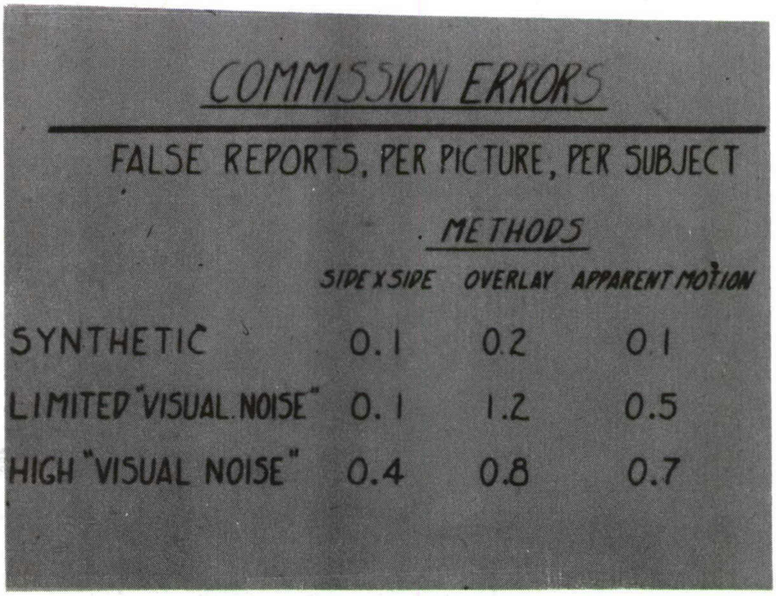


Figure 13.

The implications of these results may be exemplified by hypothesizing an orbital system which could cover all the Russian-Chinese area, about 9 million square miles, and bring back high-quality pictures. There would be about 17,010 pictures of a 9" x 9" format. If a second satellite went over and took identical pictures from the same points as the first satellite and then we were to compare them, we would first have to enlarge each picture to a usable scale, thereby increasing the number of pictures to be inspected to approximately 2,500,000. To detect the changes using the side-by-side presentation method would take about 706 man-months; and you could expect, based on the current data, to miss about two-thirds of the changes that actually existed. If the apparent motion presentation method were operationally feasible, this would reduce the task to about 202 man-months and about 7-percent error.

Our fourth experiment involved the services of nine experienced and active photo interpreters from the US Army. We compared their performance with nine of our incidental samplings of Boeing people. (Figure 14) None of the differences measured were found to be statistically significant. The difference in performance on side-by-side display method, as reflected by time scores, was not significant due to the variance of the non-professional groups. All other scores are almost identical; and actually, if these were represented as medians, hardly a mathematical difference exists. In other words, for the task of screening for change, the training and skills of the photo interpreters did not appear as any particular advantage. This supports the philosophy which we had advanced at Boeing – that we might be able to develop a display method whereby it would be possible to use large numbers of semi-trained people to do the task of screening and leave for the few highly skilled photo interpreters the task of interpreting what the changes implied. This would certainly constitute a more efficient utilization of limited numbers of skilled photo interpreters.

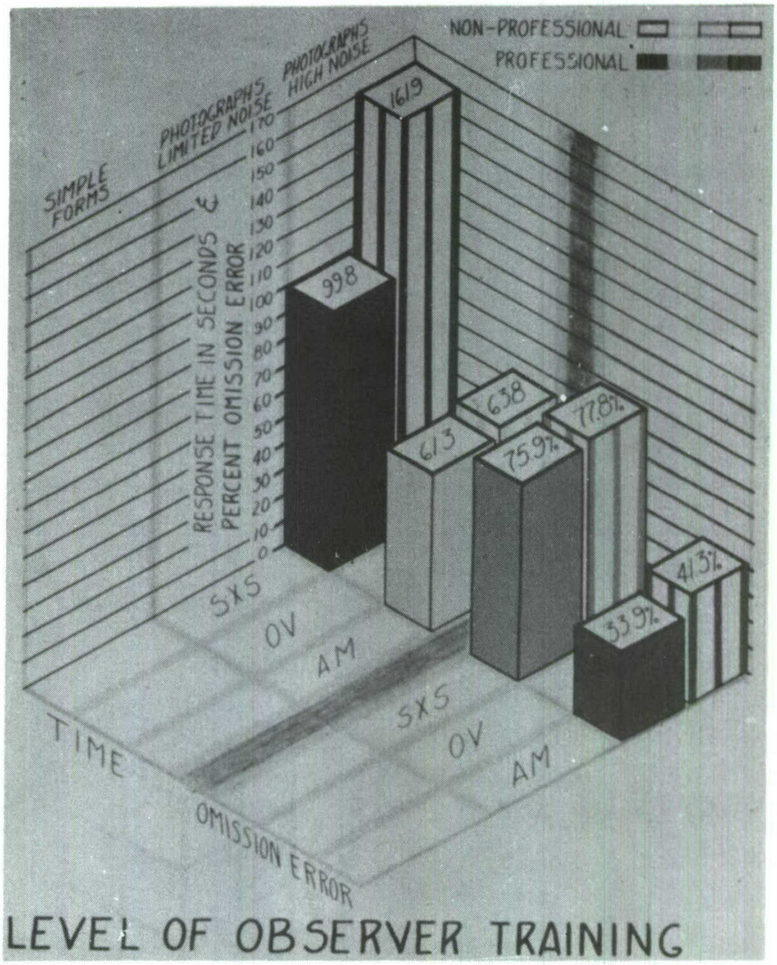


Figure 14.

CURRENT RESEARCH EFFORT

Dr. Klingberg:

We have reviewed for you some of the relevant research conducted at The Boeing Company prior to the awarding of the current contract with RADC. I have the pleasure of extending this review to include the present program plan and the performance data collected thus far. Since the research we are now conducting consists of three distinct phases, I will briefly present the basic experimental designs of each experiment and then go into specific procedural details before presenting the data.

The first phase of the program is a direct extension of the latter work which utilized the "high visual noise" aerial photographs. You will remember that this material represented comparative coverage taken from disparate points in space (58 degrees) onto which artistically introduced changes in number, position and size varying from zero to three per photo were added. In this phase we are again interested in evaluating the experimental presentation technique which we call "apparent motion," as compared to performance achieved with the conventional side-by-side presentation. We are also investigating the influence of reduced contrast on interpreter performance and have included three levels of relative contrast – 20, 40 and 60 percent. Since it has also become possible to obtain a sample of subjects who have had photo interpretation experience, we are making a comparison between performance by experienced versus inexperienced interpreters.

In our previous research we had been concerned primarily with the screening function – how quickly and accurately can observers detect the areas of change. Since this constitutes only a small portion of the information-extraction process, we are now requiring subjects to also specify the type of change which has occurred (i.e., a change in number, size or position) and the object(s) changed (i.e., oil tank, warehouse, ship, ground spoil, etc.). The specific details of the presentation and response systems which allow us to get quantitative data of time, accuracy and completeness will be discussed later.

Figure 15 shows a schematic representation of the independent and dependent variables selected for investigation in phase one

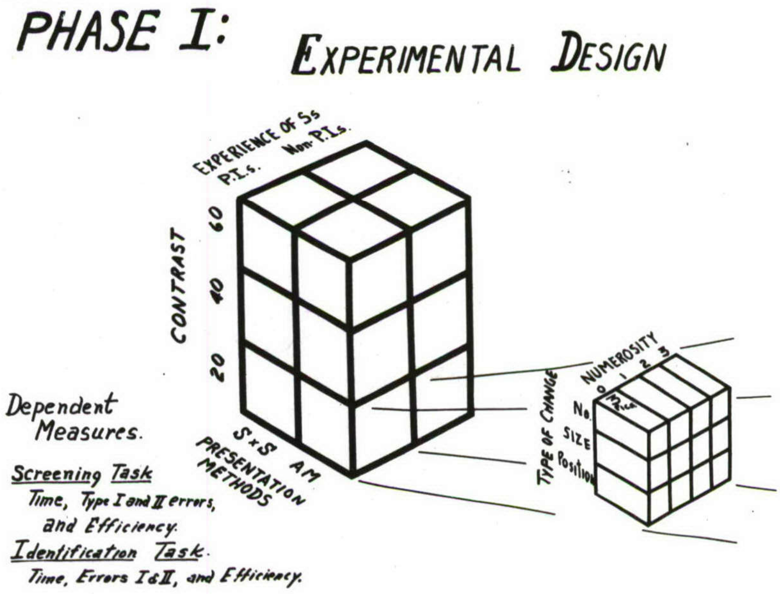


Figure 15.

The second phase of the program involves the utilization of comparative-cover photography obtained from a variety of collection and storage agencies. Among these sources are ACIC, Department of Agriculture, Coast and Geodetic Survey and the Air Force. The search for and acquisition of this photography has turned out to be a very frustrating and tedious undertaking; but with the invaluable assistance of many cooperative and understanding individuals, some of whom are in the audience today, it now seems that we may be able eventually to procure sufficient imagery to meet our requirements.

As in phase one we will again be presenting these comparative pairs in the conventional side-by-side manner as well as with the apparent-motion system. We are also interested in investigating the influence of photo scale in this phase of the research and plan to include photography representing three scale values between 1:5,000 and 1:50,000.

Two subject-population samples will be utilized, one representing naive and one representing experienced interpreters.

The photography which will finally be selected will contain samples of three major types of background: rural landscapes, suburban development areas and industrial-military configurations. Prior to presentation, the imagery will be systematically degraded by photographically varying the acutance (edge gradient) in three discrete steps.

The dependent measures will again include the accuracy, completeness and speed of target change detection and identification. As you can well imagine, the anticipated visual noise (irrelevant changes) in this photography, which is attributable to differences in taking angle, time of day (shadows), seasons of the year (foliage and ground cover), atmospheric conditions, etc., will be considerably greater than that represented in any of our previously employed stimulus material.

The third phase of the experimentation will also involve the use of the operational type of comparative-cover photography which we are able to cull from the archives in Washington and elsewhere. Although the recent missile build-up in Cuba has demonstrated a need for almost a day-by-day coverage, in some instances this is not always possible or necessary. It is conceivable that in some situations the time lapse represented between the first and second coverage of a given area may be in terms of weeks, months, or even years. In this phase we propose to investigate how performance of photo interpreters and non-photo interpreters, utilizing both the side-by-side viewing and the apparent-motion presentation system, varies as a function of the time lapse existing between the comparative coverages. Provided that the photographic sample is sufficiently large to allow adequate representation, time-lapse samples of 7, 178 and 356 days shall be selected for investigation.

The dependent measures to be obtained will again include time, accuracy and completeness.

I hope that the description I have just given you of the basic designs of the three phases is sufficient to at least introduce you to the variables which we have chosen to investigate. I would now like to turn our attention to some of the methodological and procedural details of the work that has been completed. Since the data to be reported have just recently been collected, I will not be able to report the associated significance levels for all of the findings. Wherever we have completed the analysis, I will include the appropriate statistical values; but in other situations, I can do no more than show you the slides we have prepared and let you draw your own conclusions as to their significance.

First, let me say a few words about the presentation equipment and the response system. The display system is essentially that which was illustrated earlier in our presentation. It consists basically of two optical pathways of equal length which can be illuminated either simultaneously or alternately to expose the two 4 x 4-inch comparative positive transparencies. The two optical beams are fused by a half-silvered mirror located between the viewing aperture and the imagery. By placing the two transparencies either in adjacent positions or in complete registry and selecting the appropriate lighting condition, the imagery can be viewed either as a static side-by-side display or as two pictures time-sharing the same spatial location at a rate of 1.5 cycles per second.

To introduce the three levels of contrast reduction selected for investigation in phase one, a light veil consisting of a diffused light source reflecting off of a plate-glass surface at the viewing aperture was introduced. Neutral density filters were utilized to achieve the required contrast conditions (20, 40 and 60 percent relative to that contained in the original imagery), and compensating filters were introduced into the image-projection beams to maintain a constant luminance level at the eye regardless of the contrast level presented.

In the previously mentioned research, we were concerned only with the screening function; that is, how quickly and accurately can photo interpreters detect areas of change. The present program is designed to provide additional performance data reflecting the observer's ability to identify the object(s) changed and the type of change that has transpired. This identification task, superficial as it is, required only that subjects report whether the change which they detected was one of number, size or position, and to report whether it was a building, ship, bridge, clump of trees, etc., that was changed.

In order to get independent measures of detection and identification time, a response system had to be designed which would allow separation of the two functions. As was illustrated in the earlier movie sequence, a projected 6 x 6 matrix was superimposed over the images with a light spot corresponding to the spatial location on the response keyboard. An example of what is seen by the observer at this point is shown in the next slide (Figure 16). After depressing the response key designating the area of detected change, we get a print-out on an attached digital printer which records the amount of time lapsed since the initial exposure. As he holds the key depressed, a light screen is triggered which obscures all but the designated area and the three areas immediately adjacent to it. This prevents the subject from searching further for additional areas of change before he has arrived at a decision of what has changed and how it has changed. An illustration of this condition is contained on the next slide (Figure 17). When he is ready to report his identification, he releases the key which again activates the digital printer, and provides a record of the lapsed time and drops an opaque shutter over the viewing aperture.

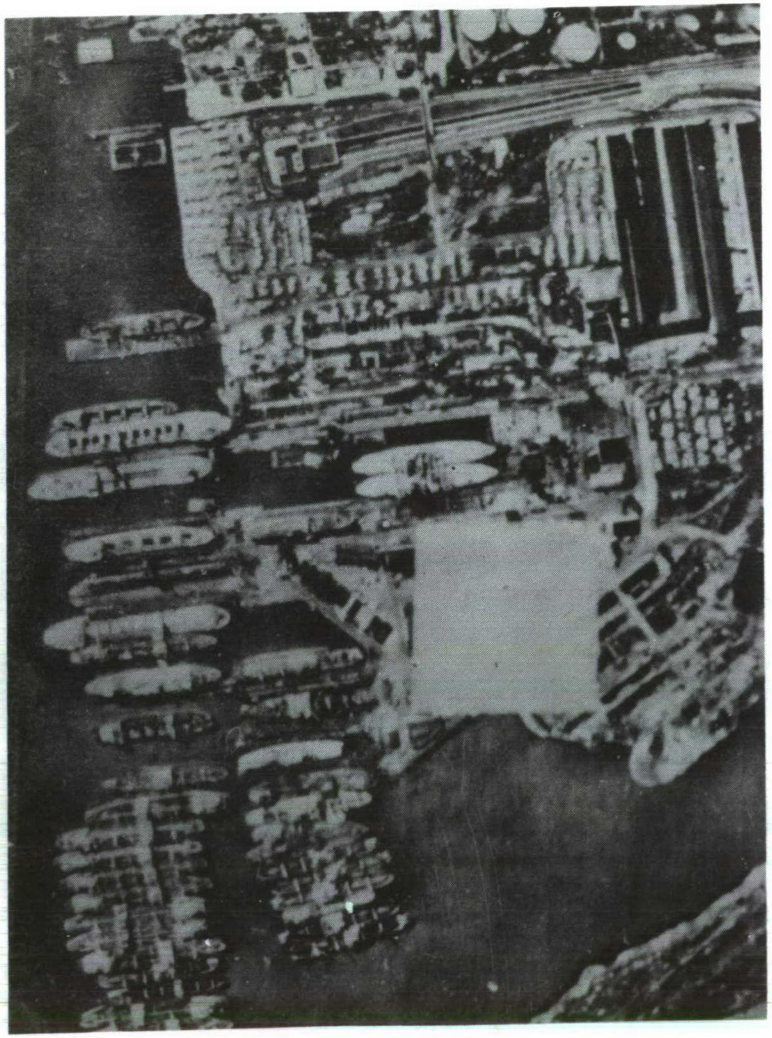


Figure 16.

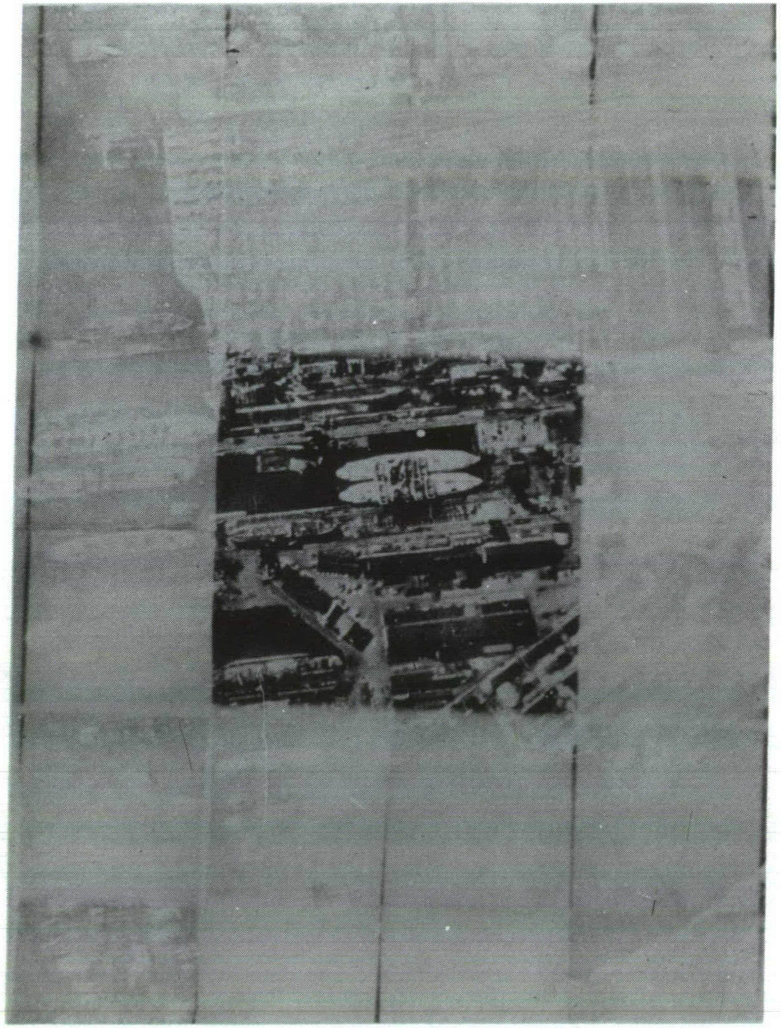


Figure 17.

He then verbalizes his identification, which is recorded by the experimenter. He repeats this operation until he feels that he has detected all of the changes on the photographs. He then punches another button which closes the shutter and gives a final readout on the digital printer to provide a measure of total time for that pair of coverages.

When we originally responded to the RFP, we proposed to utilize a fairly large sample of naive observers in this phase of the research. After the contract award, however, we placed a small item in the Boeing weekly newspaper requesting contact with anyone within the Aero-Space Division who had had photo interpretation experience. Subsequent to this initial probe, we have received over 60 responses from persons claiming photo interpretation experience. We may be stretching the term a little in classifying these individuals as photo interpreters in the military sense, since many received their training and experience in connection with geology, cartography, forestry and other allied areas of interest. On the other hand, many had also had an extensive military training and experience in photo interpretation. Since this group was comprised of individuals with such a heterogeneous background of experience, it was more appropriate to classify this group as an "experienced" group as opposed to the second group of observers used which had limited experience with aerial photography. Twelve subjects from each of these two groups served in phase one.

Let us now turn to the data thus far collected (Figure 18).

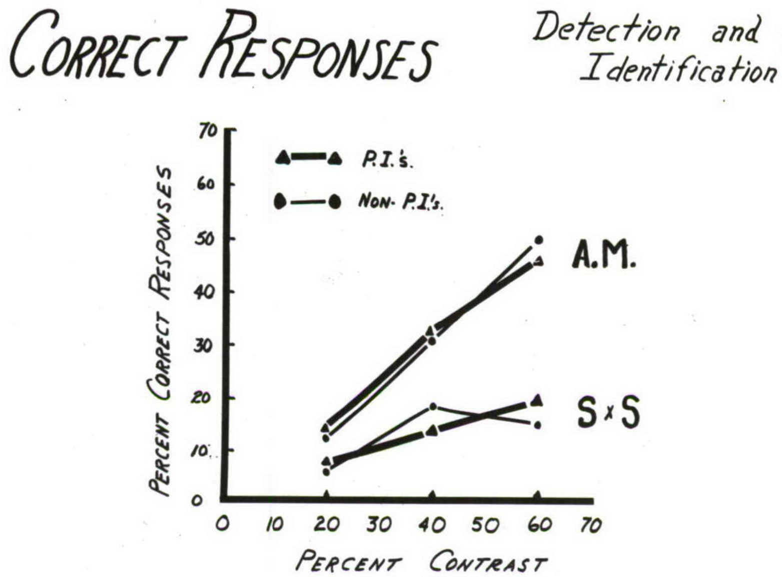


Figure 18.

The performance data in this figure are based on the application of rather severe criteria, i.e., a "correct response" means that both detection and complete identification were made. Complete identification refers to a verbal report of the type of change occurring (size, position, location, or number) and of the object being changed. We see here that there is essentially no difference between our experienced and our inexperienced groups. But, once again, we find that there is a substantial difference between the presentation techniques. The apparent-motion technique once again is significantly superior to the side-by-side technique and it appears that the difference becomes greater with improved contrast. At this point we do have the analysis of variance completed and there is a significant difference among the contrast levels and between the presentation techniques. The difference between side-by-side and apparent motion is significant at .01 and the difference as a function of contrast is significant beyond .05.

We next consider inspection time (Figure 19). Time is a very important variable when information is perishable or needed immediately, and in this situation the analysis of variance reveals that there are no differences between the photo interpreters and the naive group; but there is a significant difference between, once again, the presentation techniques. The side-by-side presentation requires significantly longer periods of time than does the apparent-motion technique. There is no significant difference here as a function of the contrast level.

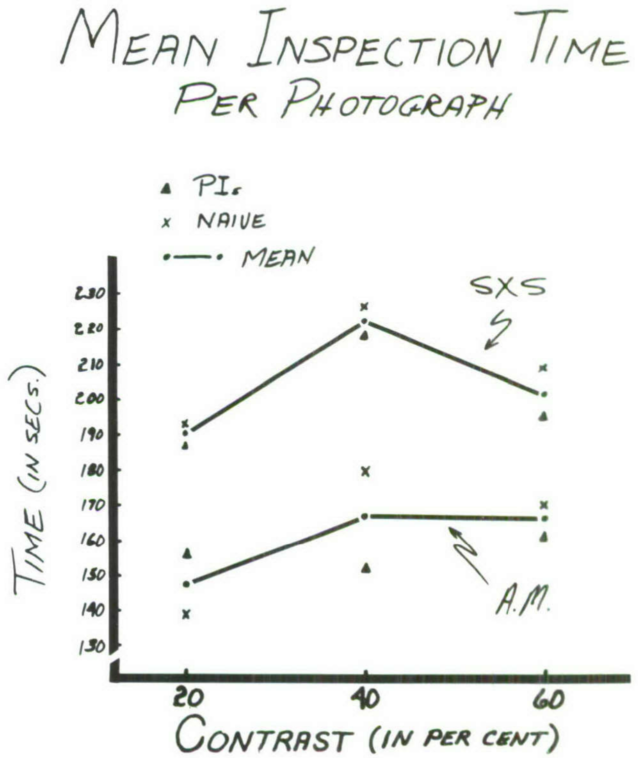


Figure 19.

When we break down correct responses into correct detections and correct identifications, we find that the apparent-motion technique once again is significantly better than the side-by-side; that it does improve with increased contrast (Figure 20). The last point on the curve represents the data from previous work with an independent group of observers. That study utilized the same stimulus materials without contrast reduction. It would appear that contrast is an important factor only when it gets below 50 percent, and that contrast changes in the upper ranges do not significantly affect performance.

As far as detection time is concerned (Figure 21), we find that there is a difference between the apparent-motion and the side-by-side techniques, where apparent motion requires less time per correct detection than does side-by-side. There, once again, is no significant difference between the two population samples, and there is no difference as a function of the contrast level.

The next graph represents an analysis of the identification task. (Figure 22) These are identifications only where a correct detection was made. What is suggested here, although this has not been tested statistically yet, is that there may be a real difference between photo interpreters and naive people, and that, under low contrast (20 percent), the experienced photo interpreters do considerably better at identifying the type of change

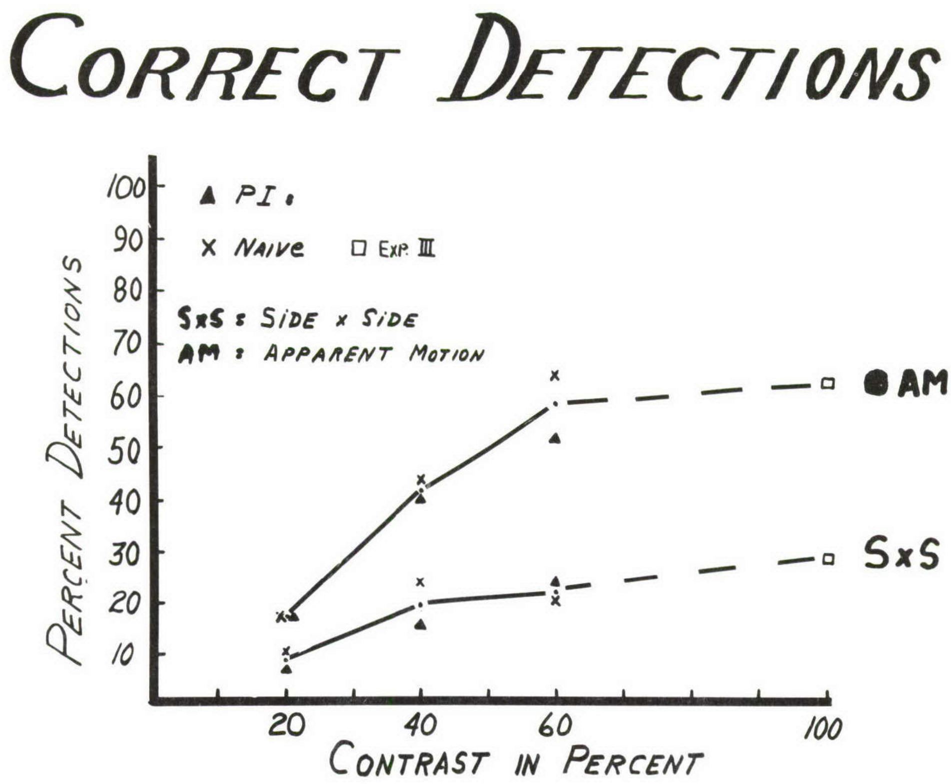


Figure 20.

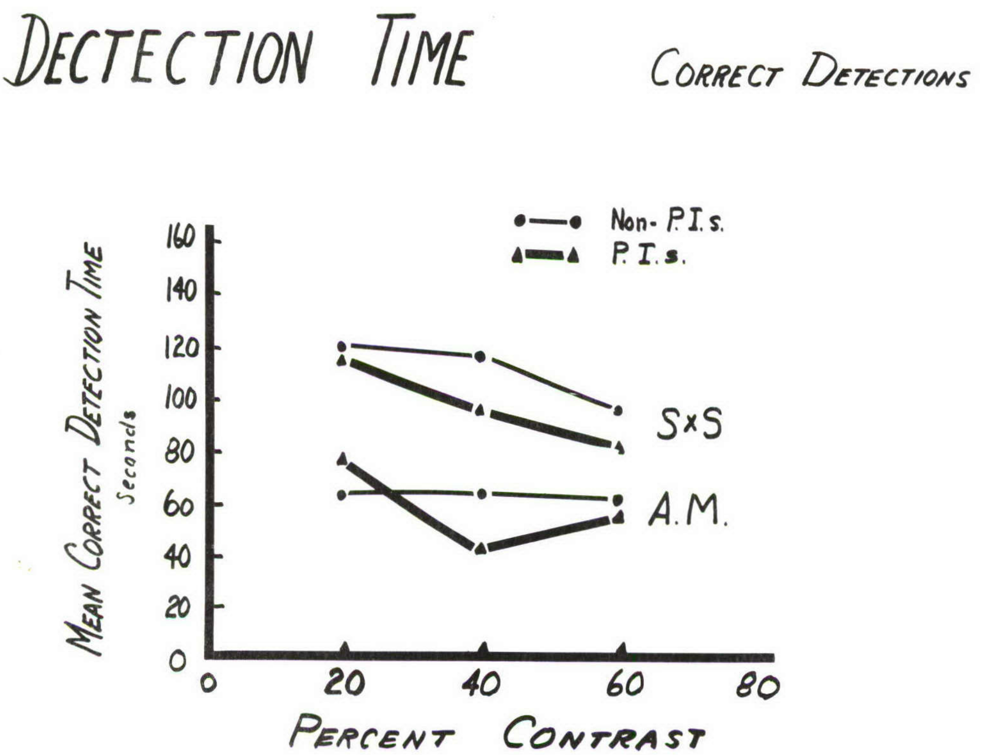


Figure 21.

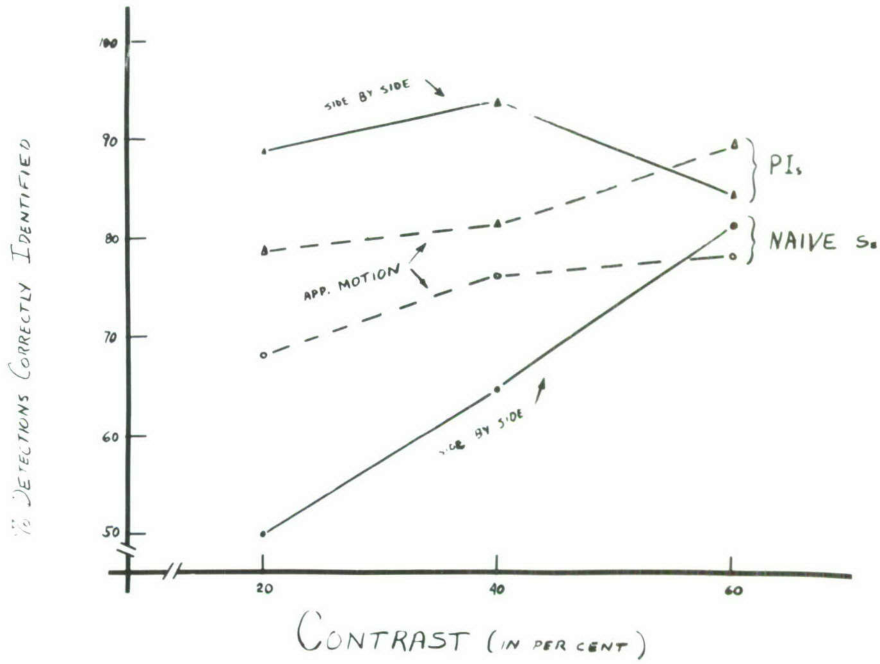
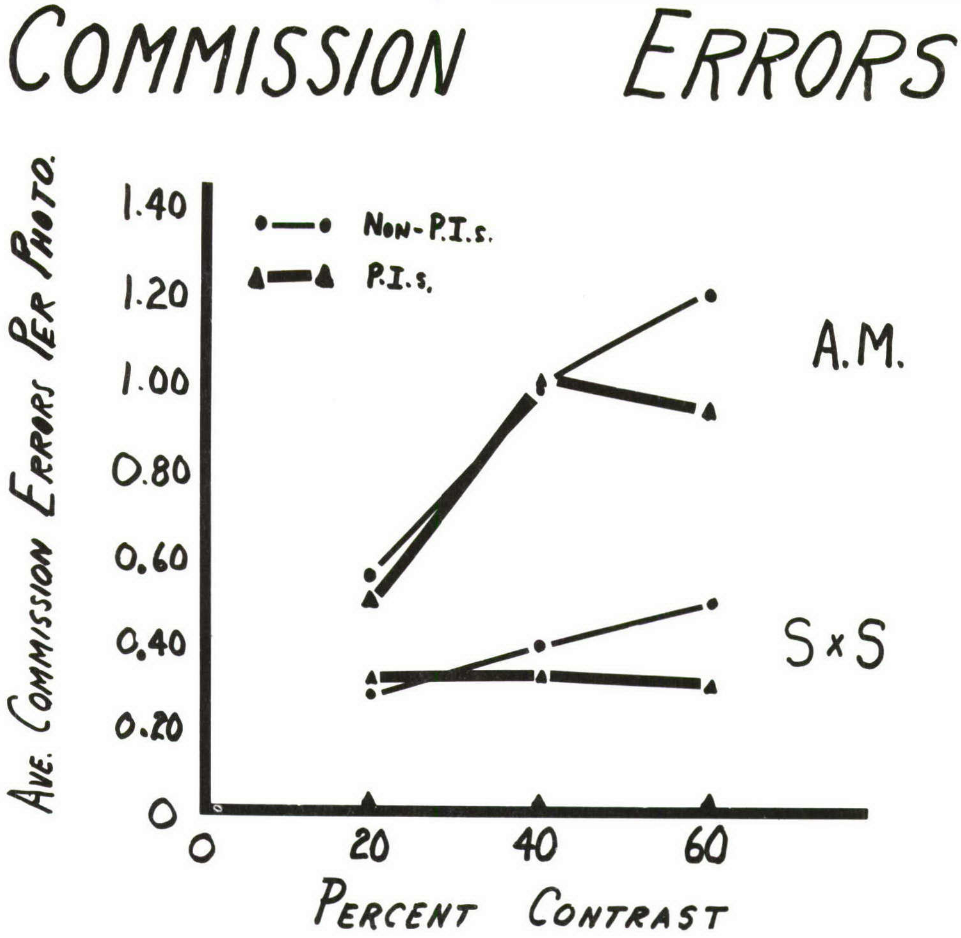


Figure 22.

which they detect. But, as contrast improves, the differences between the photo interpreter group and the non-photo interpreters seem to become less and less. This, of course, is the point at which we might expect photo interpreters to be superior. This may support our original contention that it may be possible to use naive people for the original screening of changed areas and leave the identification work to those who are specifically trained for it and could use all the time that they can get to do an adequate identification job.

The apparent-motion technique results in a significantly higher number of errors of commission than with the side-by-side technique (Figure 23). It is significantly related also to the contrast level, where the higher the contrast the larger the incidence of errors of commission. Once again, there is no difference between photo interpreters and non-photo interpreters; although at this point (60 percent contrast) we find that the photo interpreters and the non-photo interpreters are beginning to diverge and the difference is significant at the .05 level.

Insofar as we are conceptualizing this particular presentation system as a screening device, we feel that errors of commission, in the neighborhood of 1 or 1.2 errors per photograph, are more desirable than errors of omission. As a screening procedure, the material would be turned over to the photo interpreter for a detailed investigation of these areas; and, rather than overlook something that is significant due to errors of omission, it is better that a little additional time be taken by the photo interpreter to check out all areas in which changes may have occurred.



COMMISSION ERRORS
AVE. COMMISSION ERRORS PER PHOTO.
1.40
1.20
1.00
0.80
0.60
0.40
0.20
0
NON-P.I.s.
P.I.s.
A.M.
S x S
0
20
40
60
PERCENT CONTRAST

Figure 23.

GENERAL DISCUSSION

Mr. Cook (Minn.-Honeywell):

Was the 100 percent contrast which you used, normalized to one contrast ratio, and if so what kind of contrast ratio was used?

Dr. Kraft (Boeing):

The contrast ratios here are those pretty much available from CIA. We tried to maintain that ratio in producing positive transparencies from the original negatives. We do not have anything more than a measure of the brightness of the targets in contrast with the immediate surroundings, and an overall brightness of the 4x4 picture. The change in contrast was represented as 100 percent. This means that these 36 pictures and their brightnesses as represented were called 100 percent by us. We do not have those types of measure, but we do now have at Boeing, a microdensitometer which may give us the opportunity to go back and reconstruct some of these at a later time.

Mr. Meeker (Minn.-Honeywell):

I wonder how the presentation of the half-second interval between the apparent motion system was chosen?

Dr. Kraft (Boeing):

It was chosen on the basis of a very small preliminary experiment, but is what you would prefer to call an arbitrary selection. We have yet to evaluate this relative to other frequencies for these tasks. We don't know whether or not we are at the optimum part of the curve.

Dr. Klingberg (Boeing):

There is also the possibility here that in later investigations of alternation rate that we will find certain types or certain sizes of changes to be enhanced by changes in the flicker rate. It seems clear in the literature (in the abstract situation) that the Phi phenomena is facilitated by the distance between the targets and the brightness, etc. This is an area very definitely slated for investigation at Boeing to see what is optimal for what kinds of changes.

Mr. Speer (Houston-Fearless):

I was interested in whether you expect the advantages of your apparent motion effect to hold up when you start getting such gross effects as changes in season and also in sun angle; or where the finite number of changes in the photograph is not only exceeded but where changes appear everywhere and would have to be individually evaluated for their significance.

Dr. Kraft (Boeing):

If you are speaking of our hopes, that's one thing, but if you are speaking of us as experimentalists, we would like to answer that in phase two of our contract. This is what we are gathering the stimulus material for. One of the dimensions that we are trying to vary is the proportional amount of change that is possible.

Mr. Speer (Houston-Fearless):

The last question: How much trouble is it to adjust the material from the several different flights to the same scale?

Dr. Kraft (Boeing):

It is pretty much opinion at this point because we have just received this stimulus material. I do have a movie sequence, where we took two pictures from our available material and actually didn't do anything to them except change the enlargement size, put them in front of a motion camera and take eight frames of each one. They were a better match than our reproduction. What happened was that the shrinkage of the paper was in two different directions, and the difference between them shows up here greater than actually occurred in the photographs. We do not plan to try very extensive rectification, but this does look feasible.

Dr. Klingberg (Boeing):

At this point, with our presently operating first prototype hardware, all rectification would have to be made photographically. It is conceivable, however, that with refined engineering, much of the rectification can be done optically, within the system itself. We are now trying to apply this technique to the inspection of solder joints, and printed circuit cards. In this advanced prototype we have incorporated corrections for misalignment and errors along three axes, and so the machine is growing. I think that with advanced technology, we can get a system where much of the rectification can be done with the equipment rather than photographically.

Dr. Sinaiko (IDA):

I want to compliment you first; I think that this is very first rate and very exciting research. There has been one subject bothering me that we have been very consciously avoiding. It is the one of convergence of the data by your research and by others, on the lack of difference in performance between inexperienced and experienced subjects. Maybe I am opening a Pandora's box here, but I think that it's of general interest. What I would like is a general comment on the implications here. Is this an artifact of your experiment and of all the others, or is there something else here?

Dr. Kraft (Boeing):

I think that the Pandora's box that you are opening here is a whole area of research. Unfortunately, at the PI schools, the effort has been directed toward assisting in training, etc., and the whole area has been avoided because of its sensitivity, and the fact that you can't take into these the facts that we have learned through the years in terms of training, training methods, acquisition of information, etc. One thing pointed out and discussed previously is that, because we systematically changed things, we knew what would be ground true. This has never been true (according to the schools). How would one know what was on the ground in Korea, or what was on the Anzio Beach in World War II? You only got back partial information; how do you train people to do PI work unless you can give them some feedback? And feedback has always been casual or incidental, and not giving them any advantage of learning directly.

Dr. Klingberg (Boeing):

There's one additional piece of information that our data has shown that I didn't report and that's probably a function of our response system here. The PI's took longer on their interpretation than did the naive people, at 20 and 60 percent but not at 40 percent. Here, I feel that it's because of the response category that we set up. In other words, just name the object. I think that the difference in response time was because the PI's were giving us additional information. I think that this is the point at which the PI's have their maximal value, in that they can give more precise information, than can the naive people. It may be very largely a function of the nature of the task that is assigned, that either distinguishes or does not distinguish the two subject sample populations here.

Mr. Attaya (ITEK):

I think that there are two points that are relevant to Dr. Sinaiko's comments and to what you have said here. First of all I think that all of the experimenters are tending to regard all of the PI's as a homogeneous group in regard to quality, and nothing could be further from the truth. Secondly, I think that the experiments are not dealing with photo interpretation; they are dealing with seeing for PI, and I don't know of any experiments that are even beginning to establish the difference between naive and experienced PI's interpreting, and not seeing. But, since I don't have any experimental data to back this up, I've got to point out that this is merely an opinion.

Dr. Kraft (Boeing):

I have to agree with you. I think that these are certainly, for the PI, a whole dimension of tasks of which there are just a small sample, particularly on the difficulty continuum. We have not sampled the end where the PI's will really perform well.

SECTION IV

TEMPORAL FACTORS

RESEARCH ON THE RELATIONSHIP BETWEEN TIME AND PI PERFORMANCE

Dr. E. L. Chalmers
Mr. H. D. Lerner
Applied Psychology Corp.
Arlington, Va.

ABSTRACT

The present research program is concerned with the development of improved techniques related to temporal aspects of photo interpretation. Two primary variables (exposure time and work-rest cycles) will be studied independently, together with five secondary variables (scale, resolution, contrast, target density, and target detail). Anticipated research results will have implications for inflight interpretation, optimum work output; effective duty cycles; training and man-machine assignments.

INTRODUCTION

It is said that reviewing past history is a sure sign of old age. If this is true, so be it. Nonetheless, I cannot avoid noting the discrepancies between November 1952 and November 1962. Ten years ago, the photo interpretation picture was backed up by a limited flow of photography from Korea. The Air Force operated out of several small temporary structures on the west side of Lowry AFB. Naval PIC had its split responsibility of research and training. Manufacturers were in the business of trying to sell viewing devices and the research, if it could be called this, dealt largely with subject matter keys. For the most part this involved the development or compilation of keys although there was some research on what key formats might be most effective.

Shortly before November 1952 a directive came down from Headquarters USAF, a directive which would make most of us very envious today. In effect the directive stated that "between the expensive-to-obtain aerial photography and the much needed intelligence product, there is a human link" and you are hereby "directed to conduct research on this human link". I'd give most anything to have such a *carte blanche* to conduct research in this fashion today, but ten years ago it scared the living daylights out of us, largely because there was so little definition as to what this meant. Were we to conduct research on better stereoscopes? Were we to take some of the premature verbalization of a few psychologists who had looked at the photo interpretation picture and suggested that personality variables, (it's taken us a long time to live down that one) or job occupations (as Dr. Raben likes to refer to it "myopic geographers") make ideal PIs? It's most pleasing to be here ten years later and to see in extreme detail, how the picture has changed.

We are all aware of the fact that the imagery requirements of guidance systems, U-2 overflights, and satellites, have all enhanced the ability to acquire imagery as well as the demand for and acceptance of the products of imagery. This demand has clearly stressed the resources of the intelligence community well beyond its ability to provide personnel.

SOLUTIONS TO THE DEMAND FOR IMAGE INTERPRETATION

Clearly, there are three major solutions to this problem, one of which we tend to ignore because it does not include research. This is the command solution to recruit and train ten times as many PIs. The second solution is that of automation, to replace the human. (If the demands are far in excess of the number of warm bodies, then substitute for the warm bodies.) With the admitted bias of an experimental psychologist I would like to suggest that we will not know how to replace the human until we know how the human functions as a PI. Until then efforts at automation will be somewhat marginal. The third solution and the one which I believe most of us have been discussing for the past two days is an effort to improve the performance of the available interpreters. This solution, in turn, might be broken into three subsolutions: (a) Performance may be improved by enhancing the stimuli. Here we have the issues of photographic quality and the various types of screening processes that have been discussed. (b) Performance may also be enhanced by improving the human programming of the PI. This is training research and it involves the whole area of effective concept formation, enhanced identification abilities, optimum selection and sequencing of information in the training environment and the like. (c) The third subsolution is that of enhancing or improving performance by changing the environment. At the risk of over-Madison Avenuing this, I would like to suggest a distinction between static changes in the environment and dynamic changes in the environment. Under static changes we can include a miscellany of lighting, steroscopes, and perhaps on-the-job aids. Under dynamic changes we include the temporal factors which occur in image interpretation. One of these temporal factors deals with the problem of "exposure time," the amount of time an interpreter spends with a particular photograph. If you place "exposure times" end to end this would be called "pacing." The second temporal factor deals with "work-rest cycles." In this case, we do not refer to 8-hour duty cycles with 8 hours of sleep and 8 more of recreation. We are talking about more limited spans of time; the effect of 10- and 20-minute work periods under pressure, relieved by rest periods of varying lengths of time. These are the two primary variables under investigation at Applied Psychology Corporation. Our concerns then are exposure time or pacing on the one hand, and work-rest cycles on the other. Most of the discussion this morning will deal with the effects of exposure time on PI performance.

EXPERIMENTAL VARIABLES

The effects of either pacing or work-rest cycles are dependent upon a great number of additional factors, such as photographic quality, photographic content, experience and training of subjects, stimulus format, accessory equipment, room temperature, etc., etc. It would be most inconvenient and entirely inappropriate to attempt to throw into one massive experiment all of the possible variables which potentially influence pacing and/ or work-rest cycles. Instead, a judicious selection must be made of the variables most likely to have the greatest effect upon these time-related matters. The following choices have been made, partially in terms of contract limitations and partially as the result of a very short pilot study.

EXPERIMENT 1 VARIABLES

PERFORMANCE				
TASK	TIME			
SCREENING (SECONDS)	2	4	8	16
IDENTIFICATION (MINUTES)	1/2	1	2	4

PHOTOGRAPHIC CHARACTERISTICS			
CONTRAST	High	Medium	Low
SCALE	1:5,000	1:15,000	1:50,000
RESOLUTION	3 ft	10 ft	30 ft
TARGET DETAIL	High	--	Low
TARGET DENSITY	2 or more target types	--	1 target type only

Figure 1

The principal variable of exposure time or pacing is further divided into two categories. One deals with screening times–very short periods of time, during which our subjects will be asked to identify principal target types. As you can see these vary along a logarithmic scale from 2 to 16 seconds. The other set of four values for exposure time deals with the task of identification. By this we mean the recognition and correct identification of detail features in the photography. Again we have chosen a log scale ranging from 30 seconds to 4 minutes.

The secondary variables which may be expected to affect these primary variables of exposure time (and work-rest cycles) are of two types. There are the variables of photographic quality under which we have included contrast, scale and resolution, and the variables of photographic content, - in particular target detail and target density. Permit me to describe each of these at greater length. Contrast has been divided into three levels. We have had no carefully worked out formulae to apply for selection of these levels. The three levels represent what in the opinion of our subcontractor is optimum contrast, under the label of "high", "medium" contrast, and finally "low" contrast. Hopefully, the latter will be some value which is representative of the poorer contrasts with which interpreters must work. The scale levels of the photography range from a maximum ("high") value of 1 to 5,000, through a "medium" level of 1 to 15,000 to a "low" value 1 to 50,000. You can see the influence of the log function here, i.e., the expectation that 1 to 15,000 will result in performance closer to a middle of the total range of performance than the arithmetic mean of 1:27,500. Resolution is described in terms of ground-foot resolution, again on the basis of human judgment; 3 feet, 10 feet, and 30 feet. All of the initial photography is procured at high contrast, at a scale of 1 to 5,000 and at a 3-foot ground resolution. All of the other photographic variables are produced through manipulations in the processing of negatives.

The two items of target detail and target density refer, respectively, to the extent to which the target has a variety of unique features within it (target detail) and the extent to which two or more target categories are present in the photography (target density). We'll have an opportunity in a moment to look at some representative photographs in an effort to further define the variables of target detail and target density. I might add that "low" target detail has been defined as five or less unique features that would normally be identified in connection with the target, while "high" detail refers to 10 or more unique features normally identified in connection with the target. As you can see "high" target density refers to two or more targets, actually two or three targets since it is very difficult to find four targets in a 9 x 9 photo at a scale of 1 to 5,000. "Low" target density refers to only 1 target-type present in the photography.

PHOTOGRAPHIC TARGETS

The choice of target types results from a number of considerations. First, sufficient photography must be available at the scales and ground-foot resolutions, etc., that were indicated on the preceding slide. Second, the targets must be closely related to the AF mission. In this connection, an effort was made to include both strategic and tactical targets. Third, it was important that most of the features of these target types be identifiable even at scales of 1 to 50,000. Fourth, it was important that each target-type occupy approximately the same ground area so that we would not have to contend with an interaction between the amount of ground and the reduction of scale. Finally targets had to be selected in terms of target density and the target detail considerations mentioned earlier. A few illustrative examples might help. (See Figure 3)

This is *not* one of our photographs. However, it is illustrative of what might be regarded as a high density, high detail target. The high density derives from the fact that in addition to the railroad marshalling yard in the lower right hand corner, we have steel, PHOTOGRAPHY - Experiment 1

TARGET TYPES (Nine of Each)	
Aircraft	Military Depots
Airfields	Power Plants
Harbors	Railway Facilities
Metal Processing Industries	Ships

OTHER CHARACTERISTICS

NUMBER	DETAIL	DENSITY
18	High	High
18	High	Low
18	Low	High
18	Low	Low
72		

Figure 2

coke, and, I trust, iron industries; i.e., metal processing industry and railroad facilities. The "high" detail derives from the fact that among the various instances of steel industries this particular photograph contains a greater variety of the features characteristic of steel industries. Even the railroad facilities have a round house, classification yards, etc. This is a larger number of details than would be observed in an average example of railroad facilities. As you can see, this is purely a judgmental proposition and accordingly imprecise. (See Figure 4.)

This is more typical of the photography that will be employed. It could be described as harbor facilities, of "high" density and "low" detail; "high" density because the harbor facilities are directly connected with extensive marshalling yards or railroad facilities and "low" detail because there are very few uniquely different features in connection with either the harbor facilities or the railroad marshalling yards. (See Figure 5.)

This is an example of the target category of naval vessels of "low" density and "high" detail: "low" density because there is essentially nothing here except Naval vessels. (I have to add that we do not regard vessels in close proximity to harbor facilities as being of "high" density. We've assumed that harbor facilities are almost essential to a collection of ships.) The "high" detail derives from the fact that there

Figure 3

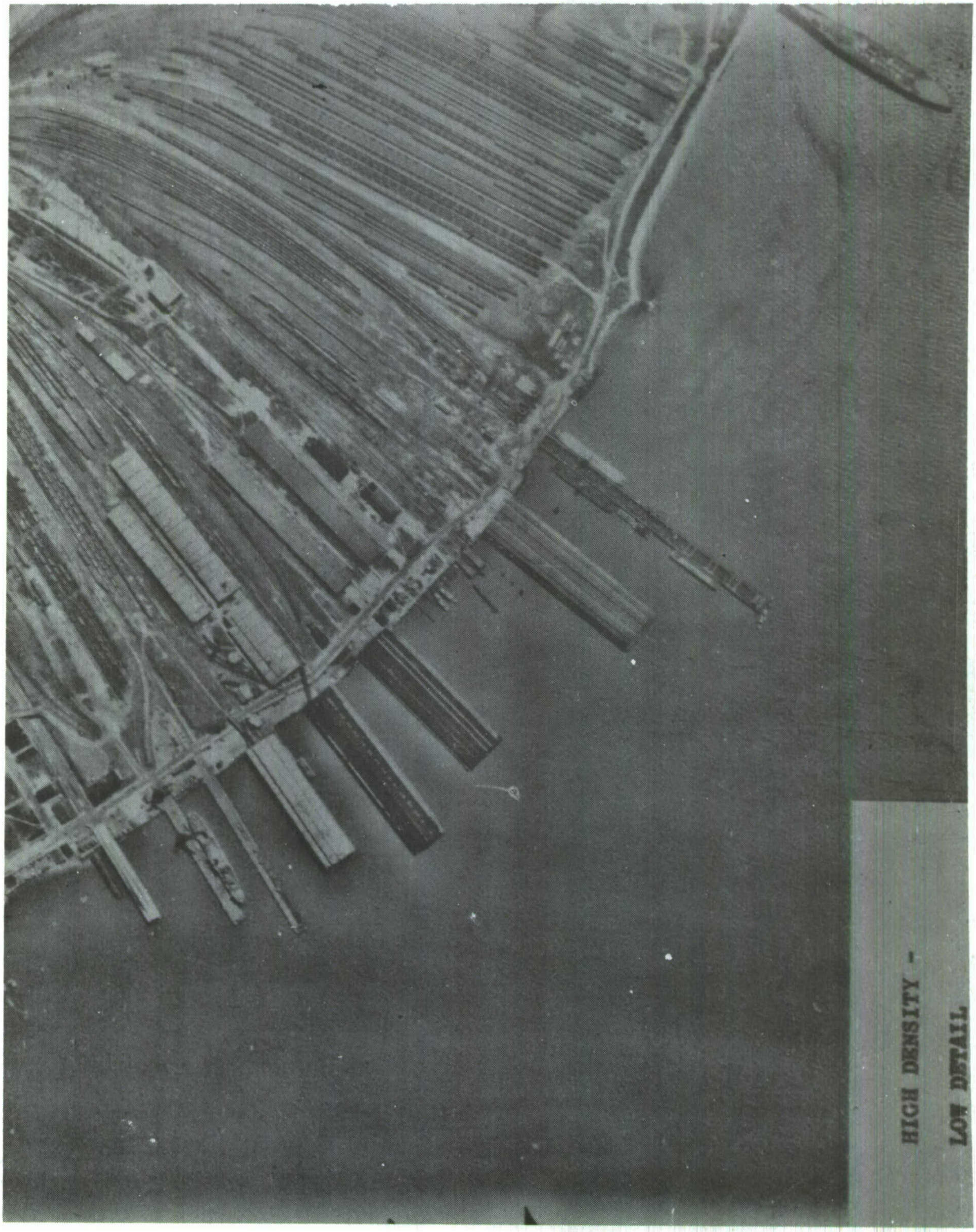


Figure 4

Figure 5

are carriers, cruisers, destroyers, and a wide variety of different types of vessels each of which would be identified by our subjects.

Finally, we have a superb example of a "low" density, "low" detail photograph. Presumably only one target type is present, airfields, and as you can see, apart from the runways, and one lone aircraft, the number of features or details are at a minimum.

APPARATUS

The next few figures show our apparatus. Essentially it consists of a conventional light table, modified along several lines, and a timer console. The film advance is remote-controlled via a console operated by the experimenter. The same console permits the experimenter to illuminate the stimulus material after it is centered in the frame (1). The 9 ½ inche positive transparencies move from the feed reel (6) over the translucent glass viewing area to the motor driven take-up spool, (3). When the transparency is centered over the viewing area, it is stopped automatically by a trip mechanism. The experimenter then presses a button on his console which simultaneously activates the timer and the lights. A dimmer control (4) allows the interpreter to adjust the light intensity to his personal requirements. There is a fan to dissipate heat. An electric brake to the film drive motor prevents the film from coasting after the transparency has been centered. Rewinding the film is done manually with the crank on the left (6). (See Figure 8).

This shows some of the inner workings of the light tables.

1 Panel containing the connectors to power, and to the timer console, and the fuse.
2 Transformer and rectifier.
3 (Arrow): Film trip switch.
4 Blower for cooling light compartment.
5 Spool drive motor; (Arrow: Control for engaging and disengaging spool from drive motor.)
6 Take-up spool
7 Exposed transparency.
8 Feed spool.

Figure 9 is the timer console showing timer covering 1 to 60 second range on the left and 60 to 240 second range on the right. Depressing center button causes film to advance to the next transparency.

SUBJECTS

It is the plan to employ trained subjects, though not experienced. There is a rather specific purpose for this. We would like the background knowledge and information these subjects possess to be as constant as possible, particularly with respect to the eight different target types. Hopefully, we will be able to use subjects from Sheppard Air Force

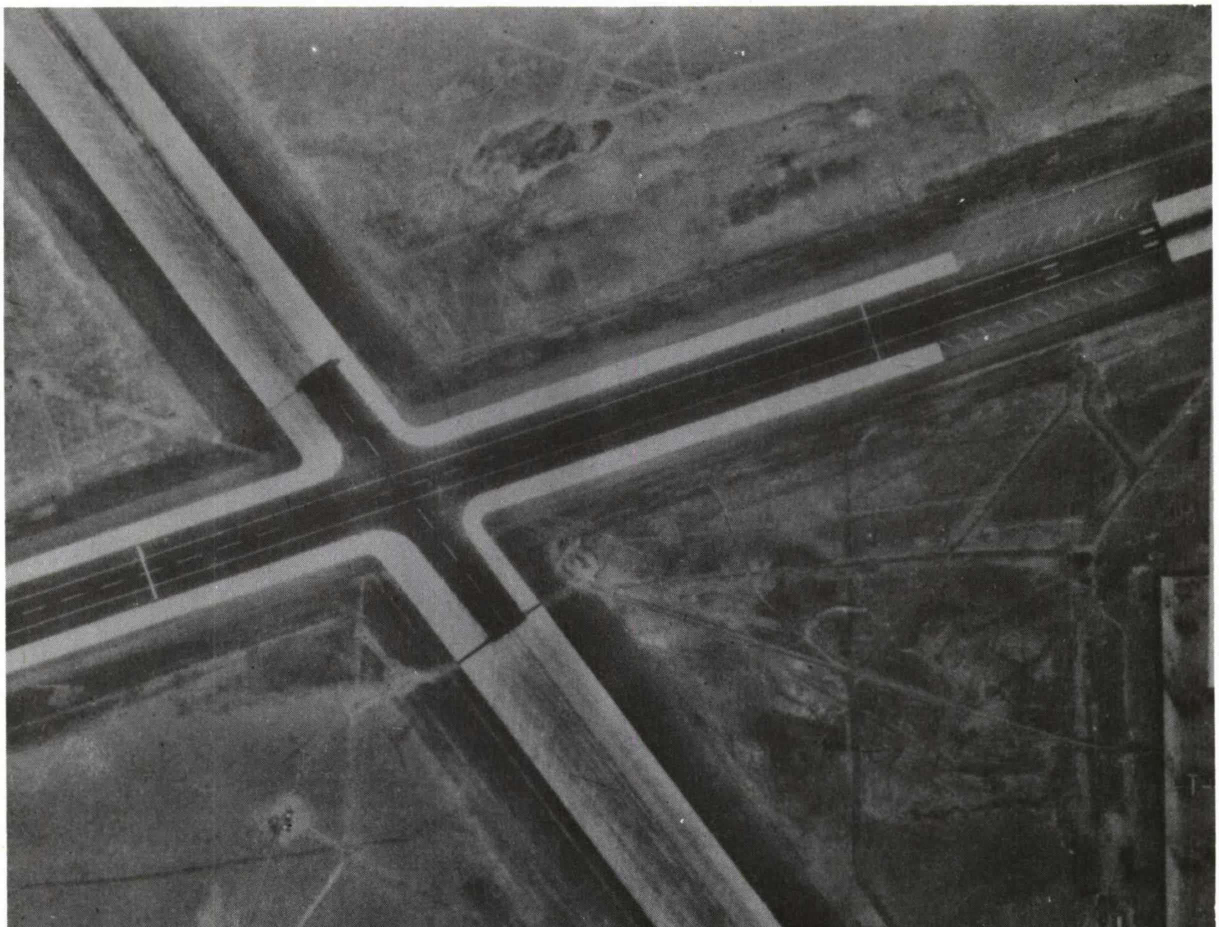


Figure 6

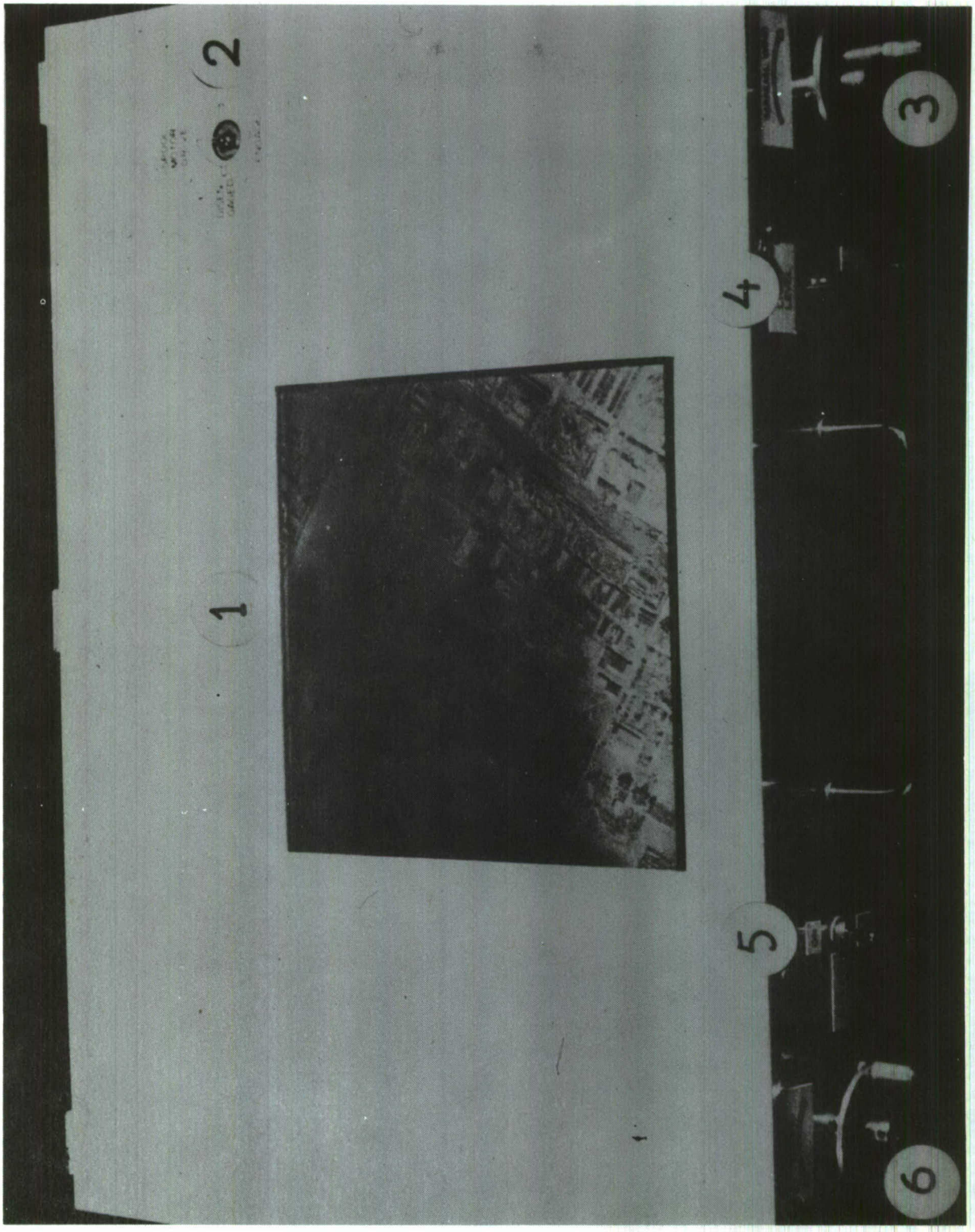


Figure 7

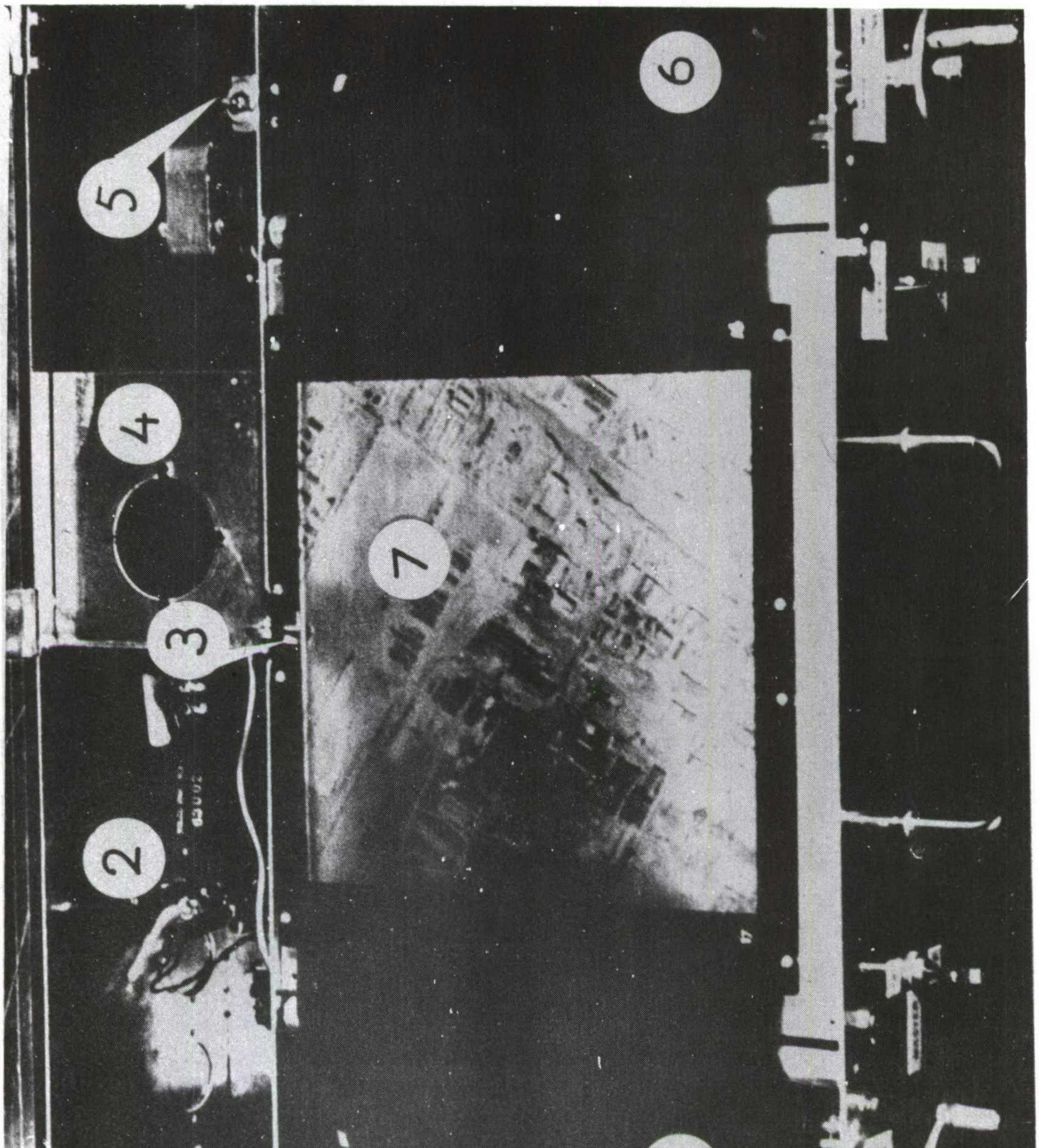


Figure 8

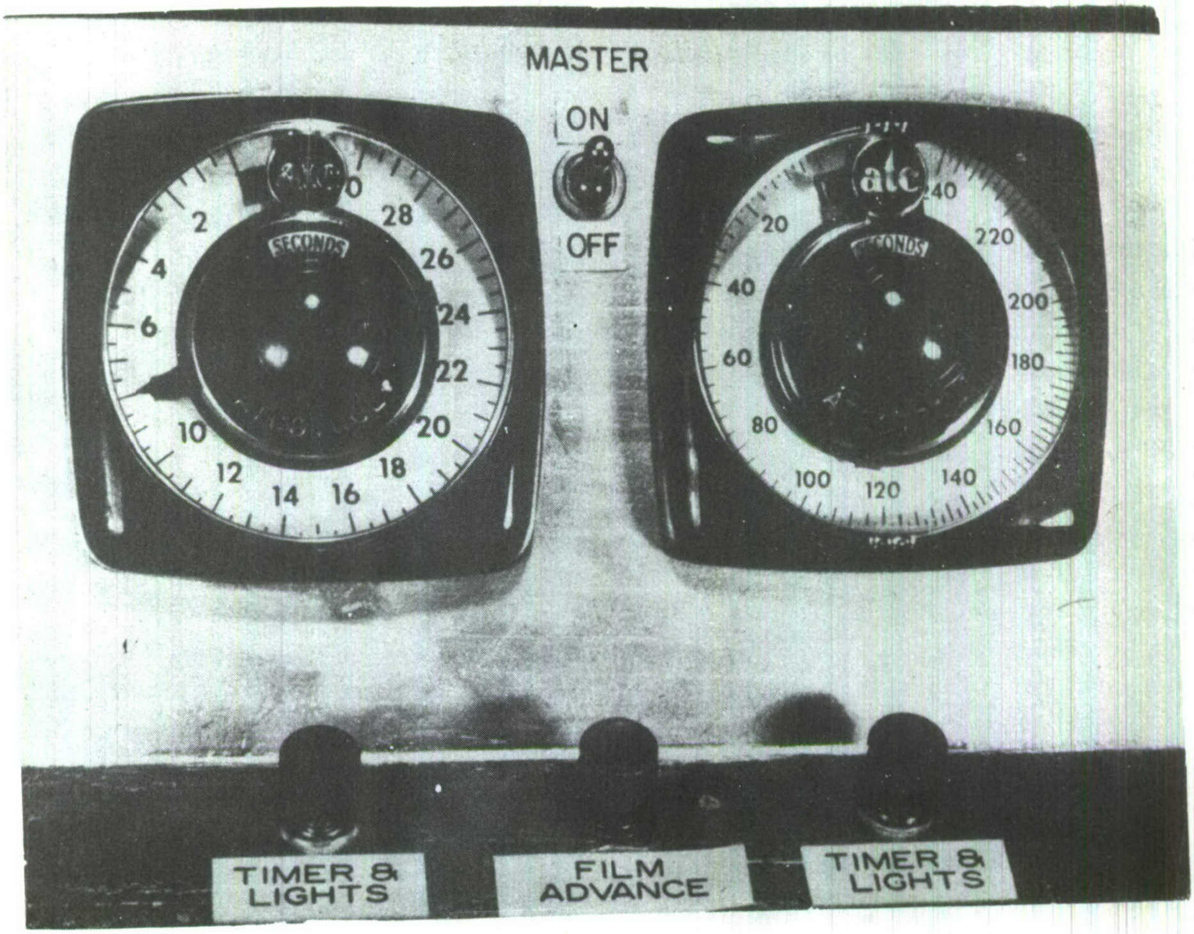


Figure 9

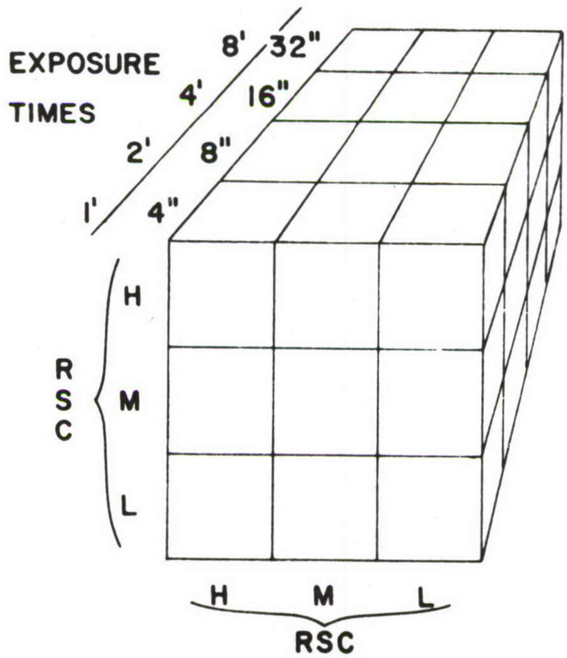
Base. We also would like to be able to run at least part of our experiment with experienced subjects. I won't embarass the gentlemen from Sheppard Air Force Base by suggesting further imposition on their experienced personnel. In response to an earlier question, asked of Dr. Kraft, I believe we will find significant differences between trained photo interpreters versus those who are both trained and experienced.

RESPONSE MEASUREMENT

As has been indicated in several other research projects reported yesterday, we intend to ask our subjects to respond to a check sheet, i.e., by checking all the target features that are present in given positive transparencies. The checklists to be provided will contain all the targets or significant target features to be encountered in the transparencies plus targets and target features not contained in the transparencies. Both the instructions and the practice periods for each subject will provide time for subjects to become thoroughly familiar with all of the targets and target features on the checklist. We do not want subjects to take a lot of time searching for appropriate responses to check. Specific instructions will attempt to prevent the checking of targets or target features which by inference should be present in the transparencies but which cannot actually be .

observed. Instructions will also require the subjects to resolve the problem of reducing "probable" or "possible" targets or target features to an absolute statement of presence or absence. For the short exposure times, only a principal target check sheet will be employed. Subjects will check target types *after* viewing the transparencies, not while viewing them. For the longer exposure times, identification checklists with targets as well as target features will be employed. In these tasks subjects may check target features both during and after their viewing of the transparency.

It is important to point out that we will have only two dependent responses to record-- number of targets or target features correctly checked, and number of targets or target features erroneously checked. All other scores, whether they are called accuracy scores, completeness scores or efficiency scores are derived from the use of various denominators with these two dependent variables. One denominator is the total number of features or targets present in the photography. When this is used as the divisor of the number of targets or lectures correctly checked, we have a "completeness" score. Presumably, if it is divided by the number incorrectly checked, we have an "inaccuracy" score. The other base commonly employed is the total number of responses made, whether they are right or wrong, divided into the number of correct responses, usually called an "accuracy" score. We will work with as many of these combination scores as seems reasonable for purposes of data analysis.



Type A Design

Figure 10

EXPERIMENTAL DESIGN

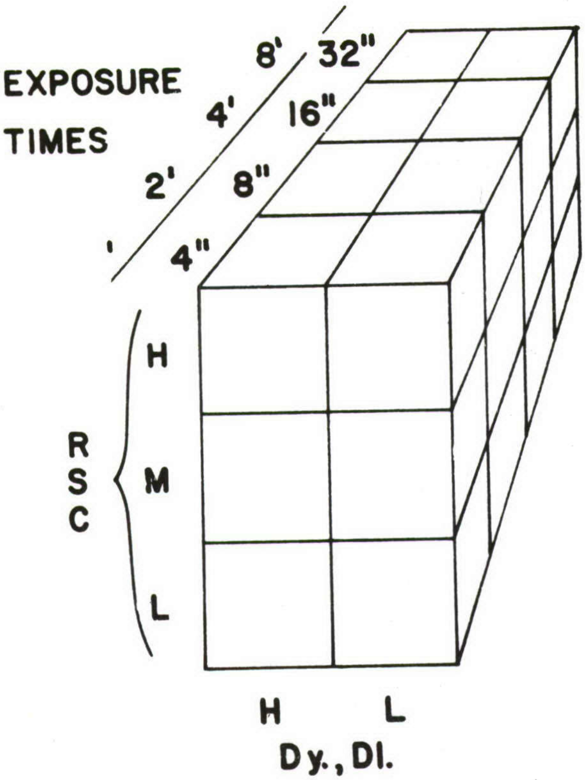
As mentioned earlier, a total of six variables are to be studied during our first experiment; the prime variable, exposure time, and the secondary variables, scale, resolution, contrast, target detail and target density. As long as we are interested in all combinations of different levels of these six variables a 6-variable factorial design could be utilized. If three levels of each of the six variables were used, 729 cells would be required. If we then introduce 8 target types we might find ourselves with a requirement for five thousand, seven hundred and some odd individually different, uniquely processed photographs and such a design would be both costly in time and in photography. We have been interested in maintaining as much realism in the task as possible and at the same time maintaining as many experimental controls as possible. As a result a series of smaller factorial designs based on all combination of two photographic variables plus exposure time will be used throughout. I should point out that we are aware of the loss of any measure of inter-and-intra-subject differences, but these must be sacrified in order to attend to the difference between the primary and secondary variables under consideration. This first slide shows a type A design, a 3x3x4 factorial in which the two-image quality variables are varied over three levels. You may read the exposure times from front to back. You may read either resolution or scale or contrast on either of the other two axis. While two of the three photographic quality variables (scale, resolution, contrast) are manipulated, the third variable remains at the medium level. "High" and "low" detail photographs are distributed equally throughout the Type A design as are "high" and "low" density photographs. Two examples of each target-type occur for each combination of the two photographic quality variables.

The Type B factorial is a 4x3x2 design which relates the exposure time (4 levels) to one of the three photographic quality variable (3 levels) to one of the two photographic content variables (2 levels). Again all content or quality variables not under investigation are held at a medium level.

The third factorial design (Type C) relates the variables of density and detail in connection with exposure times.

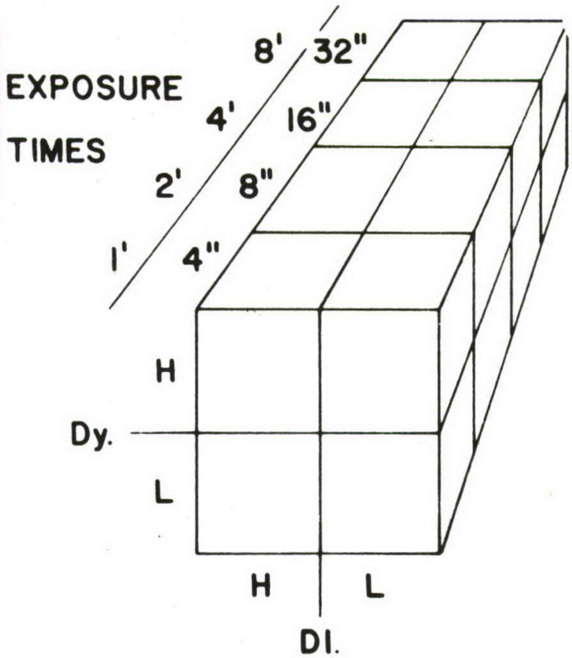
Our experimental photography will be assembled as positive transparencies on seven rolls of film, one roll for each of the three Type A and Type B designs and one for the Type C design. The order of the photographs will be the same for each of the three designs for Type A and Type B.

The total experiment will require 40 subjects, and a total of 72 photographs to be used twice by twelve subjects run on the Type A design; one set (rapid exposure times) for the screening test and one set (longer exposure times) for the target feature test. Twenty-four subjects will be run on the Type B design which employs only 48 of the 72 original photographs, and finally 4 subjects will be run on the Type C design, which requires only 32 of the original 72 photographs. The total running time for each subject is shown in the illustration.



Type B Design

Figure 11



Type C Design

Figure 12

PHOTO INTERPRETER SUBJECTS AND TASKS - EXPERIMENT 1

NUMBER OF SUBJECTS	PHOTOS	TIME (Hours)
12	72 x 2	5½
24	48 x 2	3½
4	32 x 2	2½
40		

Figure 13

Since basically each photograph is viewed only once, there is no replication, and subject variability cannot be measured. For each main effect found significant each level will be evaluated by means of t-tests.

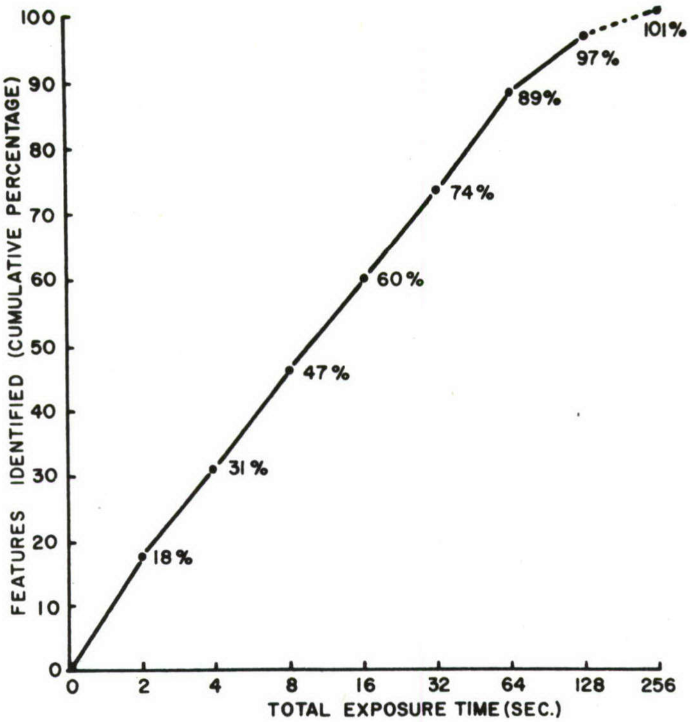
In Experiment II, the second experiment, we will be replacing the four exposure times by four work-rest cycles. Three possible types of work-rest cycles are under consideration. The first possibility involves is work-rest cycles for which there is a constant ratio between the amount of time spent working and the amount of time resting.

A second possibility is to hold the rest period at a constant level and vary the work period, thus producing a variable ratio of work to rest. Our third possibility, that of combining both of these, will complicate the analysis of data. At present we are waiting for further research, either from the first experiment or from additional preliminary studies, in order to settle upon a final decision.

PILOT STUDY

Now, I'd like to present the data collected during a pilot study. The purpose of the pilot study was to provide information needed for the choice of exposure time values for the first experiment. Four subjects with five to ten years of professional military experience as photo interpreters were used. All subjects received a brief orientation about the nature and objectives of the research and were then given specific instructions to scan the photographs for targets of military significance, either strategic or technical, as well as to identify characteristic target features. No restrictions were placed upon subjects as to whether they should report target types or target features first.

The following procedure was employed. An aerial photograph was presented at successively increasing intervals of time, namely, 2, 4, 8, 16, 32, etc., seconds until the subject reported that he had exhausted all relevant information. Subjects reported their findings verbally, during and after each stimulus presentation on both the target types and features they identified in the photography. These responses were recorded by the experimenter. Thirty-five aerial photographs were employed, of which only 26 were eventually scored. The remaining 9 were dropped because they contained redundant material. Of the 26 scored photographs, there were 13 different examples of railroad facilities, 7 different examples of ports and harbors, and 6 examples of thermo-power plants. In addition to these targets, 9 of the railroad facility photographs contained one additional target, either an industry, a harbor, or an inland waterway; and 2 contained 2 additional targets. Of the 7 photographs containing harbor facilities, 5 contained one additional target, either an industry, a military installation, or a railroad facility. All photographs were given to each subject in the same order. Neither visual aids nor stereo viewing was employed. Ground truth was established by the fifth interpreter and one of the principal investigators with prior training and experience in photo interpretation. Subjects' responses were graded with respect to ground truth and divided into target identifications and target feature identifications. Some of the results are presented in the illustration.



CUMULATIVE FEATURE IDENTIFICATION AS A FUNCTION OF EXPOSURE TIME (PILOT DATA)

Figure 14

These are rather clean data for just four subjects. They show the cumulative percent of total features identified, as a function of exposure time. At least up to a minute or possibly two minutes, this is essentially a log function. Eventually it begins to break away from the log function consonant with the fact that interpreters are not capable of finding an infinite quantity of information from a finite photograph. We might speculate about a "saturation" point. Presumably if our subjects were allowed to go to a full two-and-a-half hours they would have completely exhausted the identification of unique objects in the photography. These, then, are some of the data which provided us with our suggested exposure times for the first experiment, the times that ranged from seconds to minutes.

This, in brief, is the research in which we are presently engaged. We hope to identify which of the time variables and which of the photographic variables affect performance. We hope further to identify optimum levels of performance for given qualities of photography when viewed under various time pressures, and eventually to relate our research to such operational questions as how fast can a photo interpreter process photography either as a function of command decision or as a function of automatic processing of photography.

For a given set of photographs, with given content variables and quality variables, it ought to be possible to predict how rapidly interpreters can work, and at what level of efficiency. We hope this will have implication for task allocation and manning requirements in photo interpretation groups; where the number of personnel required to handle a particular type of task, as well as the types of tasks that can be meaningly broken down in terms of these times variables, would readily be determined from the results of our research. Finally it was noted that this does have implication for real time reconnaissance by in-flight observers whether it be overflights or manned satellites. We would also hope that the research would have implications for further research. For example, while we have not built this into the present study, it might be of interest to determine the extent to which various conditioned pacing techniques can be built into the interpreter, that is, to the extent to which the interpreter could be trained to different levels of productivity. It might also be profitable to measure inter-subject differences in response to these various time pressures with possible implications for selection.

Finally, it should be noted that research about the effects of time pressure upon types of photo interpretation activity other than feature recognition calls for considerable amounts of subsequent research. What, for example, will be the result of time pressure and of different work-rest cycles, when the task is one of detailed photo interpretation as opposed to feature of target recognition?

GENERAL DISCUSSION

Mr. Speer (Houston Fearless):

When I was going through college, I took a course which at that time was called "remedial reading", where emphasis was placed on speeding up one's reading processes far beyond that of the average person graduating from high school. As an example: the average person reads at an average of 200 to 250 words per minute, and this rate does not significantly increase through college and may even decrease. They were increasing the speed up to about 750 words per minute, and in some cases up to about 1200 words per minute, and with increased accuracy. This kind of training would have vast impact on the PI industry, and I wonder if any of your experiments have been designed to note whether, with successive training under time pressure, prolonged use of time-pressure cycle would show an increased acquisition rate?

Dr. Chalmers (Applied Psych. Corp.):

I'm afraid that the only marginal notation that could be made from the present design would be to note changes in performance over four continuous hour periods, rather than the effects of repeated training sessions with the materials. That's more likely to be a continuation of the present research rather than a part of the present program.

Mr. Cook (Minn.-Honeywell):

In your preceding experiments, if you picked any one of your target features at random and brought a fresh man to study the features, would it take more time to identify one target than another?

Dr. Chalmers (Applied Psych. Corp.):

I'm sure it would.

Mr. Cook (Minn.-Honeywell):

If this is true (and say the man might be tempted to pick the easier ones first), in the case where he is going to pick his identifications (and he had to pick ten) might this not then tend to slight the data in that he would tend to make more identifications per unit of time at the beginning than at the end of his total cycle.

Dr. Chalmers (Applied Psych. Corp.):

I'm sure that this is one of the many factors that contribute to the logarithmic nature of this type of function. I'm sure that there are others, in terms of the physical characteristics that make one feature stand out more than another.

Mr. Cook (Minn-Honeywell):

Then you might say that the difficulty of identification would be an important factor as well as fatigue in such a graph as you have shown?

Dr. Chalmers (Applied Psych. Corp.):

Yes, the fatigue factors in so far as they will be picked up in the second experiment would be independent of this by virtue of the fact that at different times during the experiment the same amount of photography and the same exposure times would be employed; thus this effect that you are talking about would be cancelled out, and we would be talking about genuine fatigue as a result of working too long with too little rest.

Dr. Sadacca (APRO):

We just completed some preliminary work in this area and there was one finding that I thought would be of interest to you. It seems that at least in tactical photography of smaller targets, over a short time, that the subjects would develop a "set", in that when they had several targets to look for, they would begin, first, to look for one particular kind of target. (They more or less just selected what they would look for.) If possible, you might want to do something about either controlling or determining whether you are getting such effects with this larger scale imagery.

Dr. Chalmers (Applied Psych. Corp.):

We noted a similar sequential effect in a pilot study. Having found the particular object or target in the photography, the subjects were clearly set for the same or similar objects on the subsequent photography. We have built into the factorial design of the experiments each type of target, a scene from each of the eight types of targets, and each type of target detail and target density within each of the block design. Hopefully, this may cancel out the type of effect you mentioned.

Capt. Hauser (Dept. of Army):

I realize that this time versus accuracy has been conducted in a lot of these experiments, and I'm not sure what we are going to prove with these experiments. A lot is needed in the area of new techniques, but the image interpreter is faced with a job of time. The best way to prove accuracy versus time, is combat. I notice that a lot of our emphasis is on the aerial photographic work, a field in which we have had years of experience, but two new additions of infrared and radar, do require newer techniques which may or may not apply to photographic interpretation. I think that we should apply a lot of attention toward these because they do provide us with some real problems for further investigations.

D.r. Chalmers (Applied Psych. Corp.):

I quite agree with the second part, about the other types if imagery and the extent to which one can successfully generalize from aerial photography to radar and infrared imagery. In respect to the first part, I wonder if the pending man-machine interface system type design in aerial intelligence doesn't impose the requirement to know about the work rate characteristics under given situations. If we begin to confine PI's with various types of mechanized equipment, do we not have to know what these rates are in order to know what kinds of assignments to make, what kinds of rosters to make, and the like? It's a system-design-type of problem, and it's the kind of information that feeds into the system design. Another part of the problem is that which was aptly pointed up by the gentleman from Boeing, the unceasing flow at a regular rate of images with just a fixed number of people to go on them. Presumably, there is a severe requirement to make optimum assignment of these people to these continuous tasks.

APPLICATION OF RAPID PRESENTATION TECHNIQUES IN PHOTO INTERPRETATION

Dr. R.L. Krumm, Mr. R.E. Graves, and Mr. A.J. Farina
Radio Corporation of America
Bethesda, Md.

ABSTRACT

A research program is described which attempts to develop and evaluate rapid recognition techniques for the purpose of increasing the speed with which interpreters can perform flash or real-time identification of high priority targets and data from sensor recordings. It is anticipated that optimal techniques may be a function of such image quality parameters as scale resolution, contrast and complexity as well as the target signature, per se. Target signatures, having a functional relationship to the target are being developed with consideration for several variables. The research design includes repeated presentations of stimulus material at decreasing exposure rates and under conditions involving differential reinforcement.

During recent years there has been an increasing interest in imagery interpretation by military and civilian groups. The "over-flights" provided quantities of intelligence data that required careful processing and minute study. Recent demonstrations of the feasibility of inspection satellites indicate that during the next decade an overwhelming mass of pictorial data will be provided which will require similar careful analysis.

It is fortunate that development of techniques to aid in accurate and rapid assessment are now receiving the concentrated attention that is required for them to keep pace with developments of multi-sensor equipment for rapid collection of intelligence data. Concurrent with the increased attention being devoted to interpretation methods, there is an increasing interest in what might be termed the "basics" of interpretation. These include such problem areas as: optimum search patterns, how best to use multiple sensor information, and optimum techniques of displaying imagery to the interpreter.

The Radio Corporation of America is investigating a central aspect of the interpretation problem. Specifically, the research is designed to develop and evaluate rapid recognition techniques for the purpose of increasing the speed with which interpreters can perform flash or real time identification of high priority targets and activities from reconnaissance sensor recordings.

A variety of target family areas is being employed, including: airfields; missile facilities; rail, water, and road transportation; P.O.L. facilities; military areas; electronic installations; industrial area; shipping; power; and atomic facilities.

The proposed study is concerned with the development and evaluation of techniques of rapid detection and identification. It is anticipated that optimal techniques may be a function not only of target types, but also of certain image quality parameters, such as scale resolution, contrast and complexity.

In general, the training approach emphasizes target signatures of specified military target areas, and recognition of these signatures under conditions of successive degradation of image quality and time available. It is anticipated that extremely brief response times may be achieved under certain specified conditions of image quality and type of interpreter function.

There are two major problems involved in training personnel for rapid interpretation of photographic materials. The first lies in providing for the acquisition of knowledge concerning the identifying characteristics of targets of significance (their "signatures"), such that immediate recognition or identification of targets belonging to a particular class may be made regardless of the background or setting against which they appear. Identifying characteristics may be defined by typical size relations, form or configuration, and relations among object features of a complex. The second problem, basic to all interpretation tasks, is the acquisition of efficient, effective scanning habits which will enable maximum information processing within a minimum exposure time.

In studies of visual search, a common observation is that free search scanning habits are frequently inefficient. Enoch (1960), for example, has shown that when observers are required to locate a target embedded in a complex background, scanning occurs in two stages. There is first a typical orientation search, in which the eye makes a brief excursion of the total display, followed by a second, more specific stage during which fixations are generally concentrated within a limited portion of the display. The typical observer fixates in the center of the display in this latter stage, disregarding the peripheral areas. This tendency has been found to be independent of the size of the display, the quality and content of the displayed image, and the time allowed for viewing. Because the targets dealt with by photo interpreters are typically small, requiring foveal viewing, natural scanning habits must be broken and the observer trained to achieve the greatest accuracy within a minimum exposure time.

It is possible that the learning of efficient scanning habits may be facilitated by aids that will establish improved scanning habits leading to consistently uniform coverage of display surfaces. Potential methods include those suggested by Townsend and Fry (1960), such as automatic scanning markers, templates which expose areas of the display systematically, or grid lines which serve to guide the eye along specific scanning paths. It should be noted that although such methods have not been demonstrated to decrease search time or notably increase accuracy in a search task, such methods might prove valuable as basic training devices prior to the training of skills specific to photo interpretation problems. In addition, specific verbal instructions might be employed to train efficient coverage.

Although specific aids and devices could prove of value, it has been argued that efficient scanning habits can be best acquired by methods which provide for progressive, systematic reduction of exposure to test material, with concommitant increases in the complexity or level of difficulty of the training materials.

The basic assumption in recognition training is that perceptual skills - accuracy, speed and inclusiveness of the perception of objects - can be improved by specialized practice. So-called perceptual training studies, as in remedial reading programs, have been summarized by E. Gibson (1958). The goal of such training is to decrease the number of fixations per line and increase the span of comprehension by controlled exposure of the materials to be viewed.

The results of psychological studies on the effectiveness of tachistoscopic methods have shown that:

1. Although performance after such training generally is better than prior performance, uncontrolled motivational factors appear to contribute heavily to certain reported gains. Studies of flash training must therefore control for temporary motivational states, which could lead to improved performance that would not necessarily reflect the superiority of the techniques as a training approach.

2. The results of brief exposure, or forced pacing techniques, show that significant improvement is strictly related to the specific materials employed. Tachistoscopic training with materials similar to those which will be employed in the actual performance task will lead to greater transfer of training, than will training with unrelated materials. It is implied that the value of perceptual training will depend in large part upon the judicious choice of representative stimulus materials.

3. When correction is given as part of knowledge of results, along with adequate repetition, the effects of perceptual training can be immediate and impressive. Differential reinforcement appears to hasten the achievement of specificity. Training with feedback in terms of knowledge of correct and incorrect responses can have lasting effects, making the corrected group superior in performance after an extended interval with no further training (Minturn and Reese, 1951).

Early USAF training in aircraft recognition consisted of rote memorization of terminology regarding shapes. This method (WEFT-wings, engine, fuselage and tail) was inefficient because it over-emphasized features which could be named to the neglect of other potentially discriminative features not easily labelled. Also, no practice was given in the act of recognition (J. Gibson, 1947).

Consequently, Renshaw and his associates, in 1942, developed a new set of techniques for the Navy which were subsequently modified and adapted by the Air Force as the "Flash System of Instant Recognition." The aim of the training, described by Gibson, was to develop instant recognition of the stimulus object. It was assumed that this could be accomplished by increasingly more rapid exposure speeds during the course of training. Greater proficiency was expected to derive from the use of very brief exposure. Flash exposure was assumed to be effective because it forced the student to perceive whole, or total forms, rather than to synthesize a number of separate features.

The standard Navy procedure of aircraft identification instruction (Luborsky, 1945) provided for early unlimited observation of the subject material while the instructor discussed recognition features. In a following series of practice sessions, the views were briefly exposed for durations from 1/50 second to 1 second, depending upon the military school, and the subject identified each view. After each such identification, the same view was exposed for three to ten seconds and reidentified and discussed. A large number of views of each plane was usually employed. Ordinarily new material was introduced at the rate of two categories per session.

In the Air Force adaptation, digit slides and slides containing varying numbers of aircraft were sometimes employed, as well as single aircraft photos, in order to increase the "span of apprehension".

Several criteria were employed for selecting the stimulus material used for training. These included the requirement that each aircraft be represented in a variety of attitudes and distances. Also, sufficient photos of each aircraft were provided so that the instructor might adequately review features without causing recognition to be based upon the features of a particular photograph.

A series of experiments to analyze the method were performed between 1943 and 1945 by the Air Force. The results showed:

1. extremely rapid flashes (briefer than 1 sec) were not necessary to attain proficiency,

2. presentation of "whole forms" was not as beneficial as training emphasizing distinguishing features, including verbal descriptions,

3. no evidence was given that "span of apprehension" training improved ability to recognize aircraft types,

4. when shape or configuration was an important discriminative feature of the class, training which provided for manual drawing of the shape was useful as a supplementary method. This conclusion has been given additional confirmation by Arnoult (1956), in a comparison of training methods for the recognition of spatial patterns similar to those displayed on radar scopes.

5. some slight advantage was found in grouping photos of similar categories, rather than randomly separating the material or presenting dissimilar photos in pairs.

Other studies, with Navy personnel (Luborsky, 1945), in general confirmed the above conclusions; (a) extremely brief exposures offer no advantage over somewhat longer exposures, (b) training with only three views of each category generalized poorly to new material of the same categories, and was therefore inefficient, and (c) rapid teaching (increasing new material per session), followed by review is the most efficient approach.

The evidence from past studies of flash recognition and from research on perceptual training provides considerable guidance as to the development of a promising approach toward training for rapid recognition and interpretation of visual stimulus materials.

1. Flash techniques, defined as training by repetition of stimulus material at decreased exposure times, should lead to improved interpretation ability. However, excessively rapid presentations appear to offer no significant advantage.

2. Rate of introduction of new material can be an important factor. Past research indicates that the trainee will acquire skills more rapidly if new categories or classes of objects are introduced at a rapid rate. (Luborsky, 1945)

3. There is some evidence that an advantage will be gained by grouping similar materials in successive presentations rather than presenting categories separately. This variable requires further study.

4. There is no advantage in presenting abstract materials as part of the training procedure. Indeed, the greatest transfer of training is expected from prior exposure to materials as nearly identical as possible to those that are to be used under actual conditions.

5. Effective training procedures require that a number of views of materials containing the target class be given; otherwise, trainees will tend to learn specific photos but will not acquire the class concept necessary to interpreting new material.

6. Differential reinforcement which provides knowledge of errors and reasons for the errors is an essential element of the training process. This knowledge could be given in the form of discussion by the instructor or by other verbal material. Analysis of the reasons for the occurrence of errors should be more effective than merely providing knowledge as to correctness or incorrectness of identifications.

In order to approach the project objective of developing techniques for rapid recognition of selected targets, and in accordance with the findings of previous research efforts, the concept of target signatures has been developed and is being refined. For the purpose of this study a target signature is defined as:

> The characteristic visually discernible features of a target that comprise the necessary and sufficient defining elements which qualify the target for membership in a specific target class and serve to separate it from other target classes.

More simply stated, target signatures are certain visual elements of an installation or target area that serve to identify it. They have, as well, a functional relationship to the target.

In developing target signatures for each of the selected target families, consideration has been given to the basic clues of photo interpretation. These include:

Size	Tone
Shape	Shadow
Pattern	Target to background contrast

Association

Of these, shape and pattern have proved to be the most valuable in developing signatures.

There are many studies that have emphasized these elements in detection and identification of targets. One study several years ago resulted in "family tree" structures for each target category. These were developed in great detail, arranging target components in tiers ranging, for example, from the broad category name - airfield - through types of airfield - military or civilian - and resulting in end items such as wind tees and auxiliary power units. The end items were components that were at the identification limit due to resolution and other factors.

Another study identified certain components of an installation as "diagnostic features" which were unique to a certain class of installation. For example, penstocks would be a diagnostic feature for a hydroelectric power plant.

A review of these concepts was very helpful in developing target signatures. Since the concern of this study is to develop *rapid* techniques, the elements, components, or diagnostic features selected for inclusion in a target signature had to be those that had some characteristic that could be instantly noted and not confused with an element within another signature. Also the target signatures have to provide rapid identification at a variety of scales ranging from very large, (1:5000) to quite small scales (1:40,000).

Originally ten target families were considered and it was hoped that for each a signature could be developed that would provide rapid, highly reliable identification, using any of the three sensors. These target categories are commonly used and are based upon the requirement for extracting pertinent intelligence information from photo reconnaissance cover. They are:

Airfields	Electronic Installations
POL	Missiles
Ports and Harbors	Atomic
Military Installations	Industry
Power Plants	Transportation

Early in the investigation it became apparent that this arrangement would have to be modified. For example, transportation in some form shows up on nearly every view that contains any target at all. For intelligence purposes and as a unique target group it is meaningful if marshalling yards, transshipment points and major transportation depots are considered. Military installations presented certain complications, particularly when they are in the middle of urban developments. It was discovered that even at rapid rates it is frequently possible to report to a more detailed level, for example, the airfield category can be identified as civil or military in most cases. It is possible to separate ports and harbors into Naval and Commercial establishments. Further modifications may be made if they appear warranted. The attempt is to prepare signatures and develop the target families to the most detailed reporting level possible.

The photo interpreter subjects used in this study receive rather intensive instruction in target signatures. Target signatures have been developed for each target family or class and for some of the subclasses. Graphic training aids in the form of slides, flip charts, etc. are being prepared for each family. These are based on a selection of "'classic" examples screened from aerial photos. Each example is presented and discussed in detail relating the visual components of the target to their functional relationship to the target family as a whole. Emphasis will be on components that will "standout" with a variety of scales and conditions.

Each of the thirty test subjects will receive about eight hours of refresher training. The purpose of the refresher training is to re-acquaint the photo interpreter test subjects with the basic skills of photo interpretation and will include the following:

1. Discussion and brief background of photo interpretation, the vertical orientation of aerial photographs with ground photographs and obliques. Illustrated with flip charts emphasizing the inherent perspective distortions.

2. Discussion of image size as related to object size. Describing the differences in image size due to varying altitudes and focal lengths. Illustrating this point with a flip chart.

3. Discussion of the three levels of interpretation.

a. Detection - stressing the separation of cultural and natural objects on the basis of regular geometric forms.

b. Identification - this is simply relating the photo image to its object counterpart on the ground.

c. Analysis - determining the function or significance of the imagery relating it to other targets or objects.

4. Discussion of the photo image characteristics that aid in identification and giving examples of each.

a. Size - relating the image size to scale and to known objects.

b. Shape - re-emphasizing that regular geometric forms indicate man-made objects. Noting that characteristic shapes have a functional and structural reason.

c. Pattern - defined here as a repetition or arrangement or man-made forms or objects.

d. Tone - describing in terms of photographic gray tones, the tendency of man-made objects to stand out from the background, being either lighter or darker as a result of the materials used in construction.

e. Shadow - discussion of shadows as being a frequent indication of the shape of an object. Noting that it is the best indication of height and relative elevation.

The objective of the rapid recognition training is to bring about a reduction in the subject's customary viewing time for detecting and identifying targets. By the time the subjects have reached this stage of training in the research program they will have undergone a certain amount of "refresher" training, and target signature training. The purpose of this earlier training will be to provide the subjects with an understanding of the elements of photo interpretation, a target vocabulary specific to this research project, and knowledge of the signatures of specified targets. In short, the subjects will have received all of the *content information* to be provided; the remaining training is concerned with increasing the speed with which this information is utilized.

The training to be outlined now is tentative and subject to modification in accordance with pilot study findings. Specific values for the variables used in the training e. g., presentation rates, number of cartridges, and degree and amount of feedback, therefore remain unspecified at this time. It should also be noted that this outline refers to the experimental group only. The control group treatment will be discussed separately.

The training can be divided into three blocks. Beginning with Block I a series of slide cartridges (each containing 36 slides) will be presented to the group. The slides within any single cartridge will be organized to contain imagery of the same scale and they will all be positive transparencies. The change in scale as a subject progresses through the series is always from large to medium to small scale. The first exposure of the group to the series of slides is made at presentation rate *A* which is the longest rate used. After each slide has been shown, the subjects record their answers (the target name) on an answer sheet during the inter-slide interval. When the answers have been recorded, the slide will be reshown and feedback regarding the correct answer will be provided in accordance with a pre-established schedule of reinforcement. In this manner the subjects proceed through each of the cartridges in the series. In addition to the brief break occurring when cartridges are changed, the subjects are given a rest period before proceeding on to the second series. Their answer sheets will be collected during this break.

The second series in Block I consists of exactly the same slides, but it will be shown at presentation rate *B* which is shorter than *A*. All other details, e.g., recording of answers, feedback, rest periods, remain the same. This procedure is repeated two more times until the subjects have finished viewing the slide series at rates *C* and *D*.

The scores achieved by the subjects during their rate *D* exposure to the slides are probably an overestimate of their ability to identify specified targets at a rapid rate. They would be an overestimate to the extent that the subjects had either learned the *sequence* of the targets within certain cartridges or were responding to *extraneous cues* in the slides. To gain a more accurate estimate of their abiltiy, a test is given as the second part of Block I.

The first portion of this test consists of a single cartridge containing 12 large, 12 medium, and 12 small scale transparencies, none of which has been seen previously by the subjects. This cartridge will be presented at rate *A* and the subjects' responses recorded as before, with the exception that no feedback will be provided. These scores, in effect, serve as an uncontaminated base line against which improvement in rapid identification ability can be measured. The second portion of this test consists of another cartridge of new slides divided equally among the three scale categories and presented at rate *D*: Thus, we are able to compare the subjects' ability to identify targets at a relatively slow rate with their ability to identify at a rapid rate.

The subjects will be given a rest period and then go on to Block II of the training. Block II is essentially a duplicate of Block I, with certain exceptions. No slides used in Block I are used in Block II. Presentation rates in Block II training and testing are limited to rates *C* and *D*. The subjects' responses are, however, recorded and reinforced.

in the same manner as in Block I. A test involving new slides is also given at the end of Block II training.

Block III consists of training only. It includes the slides which comprise the tests given earlier to the subjects. The purpose of showing the slides again is to provide the subjects with feedback on heretofore unreinforced imagery. Block III concludes the rapid recognition training for the experimental group.

The control group treatment will simply consist of providing approximately equal exposure to the same imagery used with the experimental group. No emphasis will be placed upon rapidly identifying the targets. Although their responses may be recorded, feedback will not be provided.

When all training has been completed two criterion measures will be administered. One test utilizes imagery in slide form while the other test involves an actual roll of aerial film.

The projection test will be administered to one subject at a time. The subject will be seated before the rear projection screen with a control box at his disposal. The control box consists of a spring-loaded toggle switch and a button. When the toggle switch is flipped the slide image appears on the screen. The duration of this exposure has been pre-set by the experimenter and is not under the control of the subject. (Both groups are tested with the same presentation rate). If a second or third look at a particular slide is desired, the button is pushed and the slide appears again for a similar amount of time.

The task of the subject in the Projection Test is to detect and identify any and all targets (as previously defined) which are present in the imagery. The particular slides used in this test would not be those employed during training, but would resemble the training imagery in all other respects. During the test the subject will make verbal responses which the experimenter will record. No feedback will be given.

When the slide is initially presented, the experimenter makes a checkmark and records any response made by the subject. Each repeat presentation is also checkmarked along with a record of the response. The total number of checkmarks will provide a time total for the slide, while the response notations opposite the checkmarks will give a more detailed time breakout on the level of specificity reached by the subject.

The Light Table Test will also be administered to one subject at a time. Here the subject's task is to detect and identify all target pictures in a selected roll of aerial film. The film will be a 9-inch roll (image size 9" x 9") of positive transparencies which contain a representative amount of the target types employed in the study. Each frame of the roll will be numbered sequentially to permit the subject to report the specific frame(s) in which his findings appear. A variable speed motorized light table will be used. The subject will write his identifications on a record form which would also contain his entry of the frame number. One score obtained would be the gross measure of elapsed time acquired for search of the full roll. Other types of scores consist of accuracy and completeness measures on the targets identified.

In summary, rapid recognition training essentially consists of repeated presentations of stimulus material at decreasing exposure rates under conditions involving differential reinforcement.

REFERENCE

Arnoult, M. D. *A Comparison of Training Methods in the Recognition of Spatial Patterns.* AFPTRC-TN-56-27. February 1956. (AD98197).

Bugbee, M. F. (ed.). *Manual of Photogrammetry.* Washington, D. C. American Society of Photogrammetry, 1952.

Bush, W. R., Kelley, R. B., and Donahue, V. M. *The Relationship Between Selected Physical Measures and Human Performance in a Pattern Recognition Task.* RCA Report No. TR-59-588-1, 1 May 1959.

Colwell, R. N., (ed.). *Manual of Photographic Interpretation.* Washington, D. C. American Society of Photogrammetry, 1960.

Enoch, J. M. Natural Tendencies in Visual Search of a Complex Display. *Visual Search Techniques.* Armed Forces NRC Committee on Vision. NRC Publication 712, 1960.

Fitts, P. M. *Human Information Handling in Speeded Tasks.* N. Y. IBM Research Center. Research Report RC-109, June, 1959.

Gibson, E. J. *A Survey of Research on Improvement in Perceptual Judgments as a Function of Controlled Practice and Training.* Perceptual and Motor Skills Research Laboratory, Lackland Air Force Base, Texas, Research Bulletin 53-45, November 1953 (AD24653).

Gibson, J. J. *Motion Picture Testing and Research.* Army Air Force Aviation Psychology Research Report No. 7, 1947. U. S. Government Printing Office, Washington, D. C.

Hillmann, B., Jenkins, P., and Bush, W.R. *Temporal Factors in the Accuracy of Interpreting Displays for Re-entry Flight Control.* RCA Report No. TR-60-588-5, August 1960.

Luborsky, L. Aircraft Recognition: I. The Relative Efficiency of Teaching Procedures. *J. Appl. Psychol.*, 1945, 29, 385-398.

Luborsky, L. Aircraft Recognition: II. A Study of Prognostic Tests. *J. Appl. Psychol.* 1945, 29, 449-457.

McGrath, J. J. and Harabedian, A. *Signal Detection as a Function of Intersignal Interval Duration.* Los Angeles: Human Factors Research Technical Report 9, February 1961.

Minturn, A. L. and Reese, T. W. The Effect of Differential Reinforcement on the Discrimination of Visual Number. *J. Psychol.* 1951, *31*, 201-231.

Townsend, C. A., and Fry, G. A. Automatic Scanning of Aerial Photographs. *Visual Search Techniques,* Armed Forces NRC Committee on Vision. NRC Publication 712, 1960.

Weiser, S., Shapiro, R., Krumm, R. L., and Staplefeld, W. P. *Integrated Sensor Interpretation Techniques*: Engineering Investigation and Design Study. Griffiss AFB, Rome Air Development Center, September 1961. TR 61-176 (SECRET - Title UNCLASSIFIED).

Photographic Interpretation Handbook. Army TM 30-245. Air Force AFM 200-50, April 1954.

GENERAL DISCUSSION

Dr. Sinaiko (IDA):

I have a question that is not specifically directed to RCA, but to the contractors in general. It seems that we may be missing a bet in that we have no overlap or duplication of the stimulus material that each of you is using. I wonder if one of you would comment on the possibility of using at least some imagery in common so that we might have some cross validation (or perhaps other advantages).

Mr. Sinnamon (RADC):

Perhaps I can answer that question since we are the sponsor of the program. We started the five contracts about eight months ago without any stimulus material of our own except for some libraries that we have established over the years. Because of this limitation, the various companies had to go out on their own to get the material. Some have gone to the Army, some to the CIA, some to Detachment 1 and some have waited for strip photography of tactical targets. Since the latter is SR-278 photography and is in great demand, it was impossible to give everyone the same material. Each person had to collect his own materials and use them as he thought best. One purpose of this meeting is to try to get people together so that there might be exchange of information. Now with the Boeing, Goodyear and Philco work, (I mentioned that both Philco and Goodyear are working on the change detection, and Boeing is working on change discrimination) we are trying to get a supply of comparative cover photography to supply these three firms. This is a first step in the valid comparison techniques. The Army has mentioned that they are trying to set up a storehouse of Air Force and Army stimulus material so that future contractors would have one central source. Dr. Schwartz has spearheaded this effort.

Dr. Schwartz (APRO):

In the collection of at least trisensor imagery, our conventional cameras work very efficiently. We hope to collect packages of information, which will consist of conventional photography, and multisensor material, taken simultaneously along with ground truth information. The first effort was in Panama last June, the second was at Camp Drum in September, and we anticipated going to Camp Carson in Colorado soon. This material at present will be controlled by the Signal Corps, at For Monmouth. I can't promise immediate acquisition at the present time; however, in the long run such material will be available. The problem that we all face is that the material is subject to compromise by administration to the same people over a period of time. Once we can solve that problem and get reproduction facilities and procedures for the preparation of these packets, we can make this material generally available.

Mr. Sinnamon (RADC):

Just one final word. The hope is that the stimulus material that is being used in these contracts will not be lost, but will be available for follow-on contracts next year. The most important thing this year has been production of the stimulus data. Maybe next year it will not be necessary to make new stimulus material for succeeding experiments.

Dr. Sadacca (APRO):

I wonder if I might add an appeal to these presentations. I noted that in the RCA and Boeing presentations, that there didn't seem to be an adequate representation of tactical situations. Since most repositories probably contain many more strategic types of situations, there is a natural inclination to use the latter because of its availability. However, the application of a rapid scanning technique or certain types of comparative analyses would be almost immediately in the tactical field. I think that there might be better grounds for generalizing the results if there were a more complete representation of tactical scenes.

SECTION V
INTERPRETER TRAINING

PHOTO INTERPRETATION COURSES TAUGHT AT SHEPPARD AFB, TEXAS

Captain Robert E. Millspaugh
Dept. of Intelligence Training, Sheppard AFB, Texas

ABSTRACT

Intelligence photo-radar training courses for USAF Officers and airmen taught at Sheppard Air Force Base, Texas is presented. The main course is outlined and some remarks relative to the selection and testing of students are made. The need for better selection methods of students and the requirement for more advanced training techniques with associated equipment is a continuous goal of the Department of Intelligence Training, Sheppard Air Force Base, Texas.

INTRODUCTION

The need for trained photo-radar intelligence personnel by the United States Air Force is a continuous requirement and the Department of Intelligence Training located at Sheppard Air Force Base, Texas has the mission of fulfilling this need. The Department of Intelligence Training is one department along with many others operating under the 3750th Technical Training School. This school is supported by a Technical Training Center, one of five such centers located at five different bases under the overall command of Air Training Command. All guidance and control originates at Air Training Command Headquarters located at Randolph AFB, Texas and reaches the Department through the Technical Training Center and Technical Training School.

COURSES

It would be beneficial to the members of the symposium to present some aspects of intelligence photo-radar training as conducted by the Department of Intelligence Training in order that you may have some insight on the how and why we train officers and airmen for Air Force requirements. The Department currently conducts a resident 31-week photo-radar course for officers and a resident 19-week photo-radar course for airmen. In addition, the Department offers several short courses which are tailored to requirements generated by commands in the field. For example, refresher courses for Air National Guard units, an advanced photo-radar course of 2-weeks duration for photo interpreters already in the field, an Air Tactical Intelligence Officers course for officers assigned to primarily target intelligence duties, and a 2-week Radar Prediction Course for both airmen and officers who will be involved in making radar predictions for SAC units. I might add here that the above courses pertain to photo-radar intelligence. In addition the Department also conducts a number of courses for intelligence officers and airmen. The school is now planning new joint Air Intelligence Officer and Airmen courses as required by the Department of Intelligence Agency. This requirement is a new one and will offer training for both the Navy and the Air Force. It is expected that the new courses will start in November of 1963 after final staffing and approval are obtained. In effect the new courses will consolidate previous separate air intelligence and photo-radar courses now taught.

The idea for a joint course which will satisfy Navy and Air Force requirements is not a new idea. The British have been conducting a joint course for all three services since World War II. I think both services will benefit by this program as new ideas, better understanding of each other's requirements and operating philosophies will raise photo-radar standards all around.

OFFICER COURSE

The subject matter and hours devoted to each course is as follows:

BLOCK	NUMBER AND TITLE	NO. OF HOURS	REMARKS
I	Introduction to Photo-Radar Intelligence	66	Covers basic security, PI equipment, air cameras, plotting and indexing, etc.
II	Photometrics	66	Basic math formulas, slide rule and mission planning
III	Terrain Analysis and Land Warfare	54	
IV	Aerospace Weapons Systems	72	Interpretation of aircraft, airfields, electronics, missiles, etc.
V	Surface Transportation	48	Bridges, Railroads, Ports and Harbors, etc.
VI	Basic Industries	66	Coke, Iron and Steel, Oil Industries, etc.
VII	Fabrication and Chemical Industries and Bomb Damage Assessment	78	Covers conventional and nuclear weapons
VIII	Weapons Employment Planning and Indirect Bomb Damage Assessment	78	
IX	Advance Aerospace Photo Intelligence	36	Very small scale interpretation and Automatic Data Processing
X	Photo Intelligence Reporting	42	
XI	Radar Fundamentals	78	
XII	Radar Reconnaissance	64	How to evaluate radar photography, etc.

BLOCK NUMBER AND TITLE	NO. OF HOURS	REMARKS
XIII Radar Target Prediction and Simulation	74	Makes radar prediction plates for crew use
XIV Radar Bombing	48	Throughout radar course students plan and fly 3 training missions in aircraft and 3 missions are "flown" on a simulator located in the Department
XV Radar Mission Planning	60	This is designed to give students experience in planning and presenting radar briefing to air crews. Student actually flies this mission himself.

TESTING

At the conclusion of each block of instruction, each student must pass an examination which tests his ability on the subject matter taught. A passing grade moves him into the next block of instruction. A failure is handled in accordance with his performance, both recent and past. If the student has proven to be very deficient in the subject matter he may be required to repeat the block of instruction. If his results reflect merely carelessness or mental errors he may be given so many hours of remedial instruction and then retested which, if successful, will allow him to stay with his class. If at any point along the course as a result of instructor observations and test results, the student shows poor aptitude, he may be recommended for removal from the course through board action. The Training Branch constantly monitors all tests given by the photo-radar branch to insure accuracy, objectivity, and adherence to Air Training Command training standards. A test is changed anytime it is found by the Training Branch to fall below these standards.

SELECTION OF STUDENTS

This is an area where perhaps a great deal of research is needed. The airmen students are selected for photo-radar training if they meet minimum general aptitude scores as given by tests at the Basic Indoctrination Center at Lackland Air Force Base, Texas. To this date there isn't any test given which will indicate whether or not an airman will or will not make a good photo-radar technician. The only way to find out is after he enters the school and is evaluated on the basis of his performance there. Officer students are not tested but are sent to the school on the basis of quotas which the Officer Indoctrination Center at Lackland AFB, Texas receives from Air Training Command Headquarters at Randolph. Again as in the case of airmen students, the officer is evaluated on his performance after starting the course and it is only in the student situation where any method of effectively evaluating his ability to successfully pass the course comes into play. If there were an effective pre-selection test or tests to determine whether or not a candidate will make a successful Photo Interpreter before attendance at the school it would do much to eliminate malassignments, failure rates and even remedial problems.

CONCLUSION

The Department of Intelligence Training is very much interested in the work now being done in this area and we are looking forward to seeing the results.

GENERAL DISCUSSION

Capt. Burhans (TAC):

I have noticed that in all of these presentations either there have been direct references, or at least inferences, that lead me to conclude that all of these tests have been conducted using either positive transparencies or photo prints. When we go out into the field, we are putting out reports (as near as possible) in real time and we work directly from the original negative. I think that it would certainly be of value in some of these tests if some were run using the original negative material, because we have usually neither the facilities nor the time to convert the originals to a positive transparency or a print.

Dr. MacLeod (RADC):

That's an excellent point. I believe that Dr. Saddaca has actually run tests of that type to show the relative proficiency of interpreters using the different types of formats, prints, and transparencies (both positive and negative).

SECTION VI
CONCLUDING DISCUSSION

OTHER HUMAN FACTORS PROBLEM AREAS REQUIRING STUDY

Dr. Shelton MacLeod
Rome Air Development Center

ABSTRACT

Other problem areas in addition to those discussed by the contractors are noted. These areas suggest the trend of future research programs. Discussion from the floor was invited pertinent to these and other issues and to further clarification of questions raised by the symposium.

We have reached the stage in our program where we realize that we should have at least another day to give all of you a chance to say the things that you've been wanting to say before we wind up our program. I would like to list a few areas which we have tended to neglect in our discussions. I'll just mention them and then throw the discussion open to the floor so that any of you who desire to elaborate on other ideas will have a chance to do so. Certainly one area that needs attention is the question of selection criteria for photo interpreters; development of reliable indices and test batteries which will permit us to select photo interpreters on the basis of the required personality characteristics and basic abilities. What is it that makes up an effective interpreter or, what is it that makes up an effective specialist of a certain type? This whole question needs some systematic investigation. Another very important area in the training problem, I think, has to do with the application of teaching machine techniques for instructing interpreters. Certainly I think we are faced with the problem of requirements for many more interpreters; at least, that would seem logical; and then we are faced with the great training burden; how do we get them trained in time, how do we train the numbers required? I think that the teaching machine techniques and programmed instruction procedures have indicated in other areas of educational curricula that individuals can attain levels of efficiency at rather astonishing speeds for selected types of programs, and we should inquire into their application to photo interpretation. Perhaps they have application to improving the ability to detect target signatures, to improve the problem solving capability of the interpreter or his ability to make proper associations; etc. .

Another area would have to do with the improvement of reference materials. New, improved or special-purpose keys, charts, and in the same breath we can say that we certainly need some human engineering projects to assess the quality of current interpreter equipment mensurational devices, stereoscopes, and the like, and to see if there are proficiency improvements which better fit man to his function. The potential usefulness of color as an aid to the interpreter needs to be studied. Perhaps providing this parameter in our photography will enable the interpreter to make fine discriminations which cannot be made with black and white format. I think this has been demonstrated in areas of forestry and geology; certainly we expect that it could also be applied to military situations. Interpreters must be trained in more effective search techniques. We have not talked sufficiently about the search function here and of the problem of locating targets. It would seem that interpreters could be trained to do this more efficiently or develop scanning equipment that would aid him to do this in an operational situation. We talked a little about enhancement techniques, contrast dodging, and spatial filtering; it looks good, but do we really know that this improves performance? Do we not need to perform a search which will objectively demonstrate increased proficiency on the part of interpreters when they are given the benefits of such enhancement techniques? Finally, another type of problem that requires consideration is the language the interpreter uses. If interpreters could use a standard language would this not improve the reliability of their reporting procedures? Also, if the interpreter communicates with an automatic reference storage system, an automatic library, the use of a standard language would permit him to query such a system and to provide it readily with his input (which only he can make) to retrieve information from it which will further aid him in doing the task that he has to do.

GENERAL DISCUSSION

Mr. Speer (Houston-Fearless):

You alluded to color transformations as possible ways of improving PI performance, and I have here some samples of work where the object is to convert the density level on the original black and white, to color. The idea is that the eye, like the ear, is more sensitive to frequency than it is to amplitude; and, therefore, if we were to convert amplitude information into frequency information, certain things might appear more readily. We do not suggest that this material is either final or usable for direct interpretation, but merely, in one case, you just shorten the search procedure considerably. We have an example here of the Washington area. All of the roads turned out red, the trees blue, and the houses a mixture of green and white. Interesting, in that it now allows you to find man-made objects amongst natural objects since they stand out in a widely different color. The object then is to place your color cuts in such a way as to make the small density differences into wild color differences, calling your attention to certain areas like a flag. We were not sure when we started this that we would get this happy result. Unfortunately, all of these photos suffered from registration problems. I have then here for anyone who would like to examine them afterwards as examples of what I think is a very exciting technique coming up. I also don't know how this will work on the particular kind of material that we have had in our hands. We have been hampered, as many others have, in obtaining representative material over a wide range. Further than that, I would like to mention a machine which people have alluded to from time to time. I am presently installing the image quality meter built by our company. This machine is not a desk-type instrument; it may be the eventual "voltmeter" of the photographic trade, except for its 1400 pounds and large dimensions. It has shown us that we can rapidly evaluate live photography rather than test photography in terms of such things as granularity, sharpness, and, to a limited degree, resolution.

Mr. Hansen (Nortronics):

Most of the discussion in the past two days has centered on individual performance in the interpretation business. I am wondering if anyone is working on optimization of team performance (the breakdown of the PI problem perhaps into the areas of specialization). Some implications have come up here; Boeing's work implies a breakdown in terms of screening and detail interpretation. Nortronics is relatively new in this aspect of PI, and I wonder if anyone is working from the team point of view?

Dr. Sadacca (APRO):

We have recently completed some work although we haven't analyzed the results. We have tried out ten different team methods, collecting data for fifteen performance measures over two half-day periods with about eight teams under each method. We were varying the size of the team, the amount of information played between the members and the roles of the members. As I said this data has not been analyzed, but within about six months we should have a report on it. As long as I am discussing team effort, I might as well say why we put so much emphasis on this. It seems that team methods offer one of the best procedures to get "out of a box" (this was mentioned by Dr. Roetling when he talked about the conflict of criteria). I think this is a very interesting area for research. Now what is the conflict? The conflict involves the attempt to maximize two things that are incompatible; namely, completeness and accuracy. We have seen a number of curves indicating that the longer a man looks at imagery, the more targets he will see. But, the longer that he looks, the higher the probability that any identification he makes will be erroneous. So on the one hand you are increasing your completeness, and on the other you are decreasing your accuracy. Through team efforts we hope to discover, by mutual agreement, a way by which (either through independent or cooperative analysis) you can get the advantages of both completeness and accuracy because then you would tend to agree whether or not the object was really there. I think that the efficacy of a procedure will depend on the particular situation that you have in mind. That is, if the situation is one where you have to maximize completeness, the interpreter can well use one kind of technique, whereas, if he is trying to maximize accuracy, he should be reporting at a different level. I think this relates to the whole question of the confidence level at which an interpreter makes an identification, and to the specific operational procedures given by military commander i.e., should the interpreter be emphasizing accuracy, or completeness. If he is trying to emphasize completeness, he would respond to every blob on the photograph. That of course is unacceptable, because accuracy loss would be prohibitive; whereas, if he is trying to emphasize accuracy, he will only respond to those blobs that he absolutely is certain of, and hence completeness would go down.

Mr. Hauser (Department of Army):

You are involved in work that certainly is needed in training the interpreter and in establishing the significance of his interpretations. I do want to warn you not to accept the idea that machines will replace the PI. The machine can't analyze the image in terms of what is required in intelligence. Also a lot of emphasis is being placed on screening devices which run through five hundred feet of film and pick out a photo which might divulge a target or target complex. This is also in the fringe area because there is much information potentially available; for example, in partial cloud coverage or in areas where it is difficult to establish complete significance. Any image interpreter working would like to have his "warm paws" on this. Therefore, he would be more inclined to pace the photography frame by frame until he is satisfied in his own mind that all targets have been identified and located. As far as emphasis on team effort, I think that it is only human nature to use the "buddy system", since one man's skill or competence differs from that of another. Although each man may be capable, they will always desire consultation. As to the question of how to scan imagery, I think that we can take a good look at ways to motivate the interpreter by use of moving presentations. There are many.

possible techniques, and if we can come up with any one that would satisfy most requirements, then we should stress this. Do we go diagonally, do we go horizontally, do we start with some distinguishing characteristic and go from that? I don't know the answer; it's an interesting question and one that needs exploitation.

Dr. Krumm (RCA):

I'm afraid that we may have frightened some people with our presentation. I want to point out that we certainly don't think that the PI is going to be looking at this for one second. All we are trying to do is to use a more rapid presentation rate as one of a number of possible methods to get this PI to the point where he can rapidly pick out the salient features of a display. He doesn't have to study this for thirty seconds and say this is a military target; he will pick out the distinguishing features and say that it's an airfield in the first few seconds. On the criteria measures we are using the standard 9-inch film roll. This will represent the operational situation to some extent. We are using a gross measure of elapsed time to go through the entire 400-foot roll which contains about three dozen targets. It is our hypothesis that after completing the exercise (being machine paced or speeded up as it were), this experimental group will be able to go through this roll at a faster pace than a control group that has not had the training. This may have some application to realtime interpretation, but we don't envision an operational machine that flips pictures at this fellow at half-a-second intervals.

Dr. MacLeod (RADC):

This project is not unlike the valuable type of training that the aircraft spotters received. The question is, can we apply something like this to photo interpretation? RCA is at a pretty early stage in developing an effective technique, which will have to be validated.

Dr. Roetling (CAL):

Dr. Sadacca brought up the question of conflicting criteria and I think everyone has precisely this question in mind in trying to make the experiments both realistic and yet controlled. I'd like to direct a question to Capt. Millspaugh or Capt. Burhans in terms of the operational situation of the PI. Specifically, as Bob Sadacca pointed out earlier, the PI has a conflict between accuracy and completeness; yet, obviously he resolves this conflict consistently in operation. I'm wondering, first of all; do the PI's work closely enough with the military situation to have very definite standards (perhaps in terms of how many weapons to be dispatched, etc.) that sets for them the number of targets that they're really looking for, therefore, essentially setting an accuracy criterion. Would Capt. Millspaugh, from his training experience, think that it would be possible to set up an experiment simulating the actual PI conditions in the field to produce precisely this type of set and give us the normal operational comparison of accuracy and completness without artificially picking one or the other in our test program.

Capt. Millspaugh (ATC):

Normally, a reconnaissance mission is laid on to satisfy certain goals. Either it may be to review a target that we already know about to indicate what changes are taking place; we may be just covering an area that we want to keep under surveillance (to see if anything is developing there); or there could even be strategic reconnaissance where we're concerned with an entirely new area with no prior intelligence where anything that appears on the photography may be of significance and would have to be carefully screened from beginning to end. How an outfit operates varies according to its own organization make-up. Usually you have a team of PI's under a supervisor (the supervisor essentially being a more experienced man). The degree of experience within that team is certainly going to vary. In terms of reporting reconnaissance, one individual may be given a specific task by his supervisor for a particular mission. This individual has in effect a check list, because most reports are in a sense a check list. Certain specific requirements must be met in terms of what targets are located, where they are, what their significance is, etc. Before a report is compiled, the individual may not be completely satisfied with his own results. He will not only consult with other PI's, but also may use other reference materials (for example, keys and ground information). Some jobs are fairly routine and do not require much from the average PI; some particular installations are peculiar, you've never seen them before, there isn't any target signature (to use the term) to signify what this thing is. In some cases, to identify its function you've actually got to call in experts in areas: Atomic Energy, Weapons people or people who are familiar with various other industries and can provide additional clues for target identification.

Dr. Schwartz (APRO):

I had been thinking that there are probably several categories of requirements imposed on the PI. He may examine an area for targets of opportunity (one kind of set), or in another situation he may be looking for discrete items and he will note that kind of target only.

Capt. Millspaugh (ATC):

This ties into the team concept also. After a reconnaissance mission, a supervisor will direct individuals to one area of responsibility, (to locate and report all airfields; or pick up missile sites; or military electronics; or transportation). The entire batch of prints (and this can be quite extensive) will be passed out to each group who will come up with only the targets they are responsible for. Later this is all integrated into the final report. But again, you always run into something on a roll of film that you were not looking for, and this will have to be explored. As a general rule, either you are asked to locate, identify, and report on a specific target or you may be required to go across the board and report all intelligence. This may speed up the operation, in terms of only selected targets being asked for, but, if it is the case of a surveillance search over a new area, a priority is established in terms of the objects that you want reported: for example, the first thing that we want reported in terms of time is a report on the missile installations, airfields have second priority; troop disposition or troop movement, third priority; industries, fourth; target complexes, fifth. You want them put out in that order. The biggest problem in terms of the PI, is timeliness. This has been a considerable problem for us; an aircraft flies three hours to a target area, photographs it, requires three more hours to return and then two more hours are required for processing the film before you finally get to the heart of the matter. In terms of constituting tactical intelligence, your information may be completely insignificant by that time. A particular convoy, or missile unit may have been at a particular place at the time it was photographed, but where is it now? A lot of ideas have been mentioned as to how we get detailed information in a timely manner. One idea was to have inflight processing; this hasn't worked out too well. Another idea was to have inflight processing, and give the navigator a fundamental PI training so that he could radio back what he has seen. Another idea was to have a television camera on board the aircraft, using a ground-based PI monitor, but there would be quality degradation from the monitor which may not give him enough detail.

Dr. Roetling (CAL):

May we take, as an example, troop concentrations which are difficult to see in aerial photography. In a situation like this, how is the interpreter told how doubtful he can be and still report that he thinks that there is something there. It is not the question that he can see them clearly or the time that it takes for him to search the photograph. Even when he's searched the photograph for a reasonable length of time, however, he finds some things he doubts are even targets; he thinks that they may be, but he doesn't know. Somewhere in here, before you write your report, you have to say yes, you think that they are, or no, you think that they aren't. This is a question of accuracy or completeness. Someone has to make this decision of yes or no, and in the military situation just how is this done?

Capt. Burhans (TAC):

Normally, in the case of identifying troop formations, you are not working in the real-time situation mentioned before. You have a matter of hours to put out this information. Probably this information is generated by the ground troops in the area, to the effect that the enemy seems to be building up, and perhaps has been for a week. Normally you would study any former coverage of the same area for various signatures (such as the number of new paths that have suddenly appeared in the woods or new supplies in an area, etc.). Considering all these things together, you know that the enemy is building up in this area.

Dr. Schwartz (APRO):

This is the expression of confidence level that is normally given with a report (positive, probable, or possible). In the case of troop activity, which he is not very sure about, (small differences exist between present and prior photography) he very commonly reports troop activity, vehicular activity and probable troop buildup, although he never sees the specific or discrete object itself. He sees indications of the activity.

Capt. Millspaugh (ATC):

This is something that he has to be very careful about, obviously. He is basing his interpretation on two things: mainly on the photography, but also on his own judgment. We have to be careful about being misled by artificial disposition of troops or the use of dummy airfields. A dummy airfield would make a beautiful signature, but in actuality there is nothing there. The Germans, in the last war, invested considerable effort trying to divert our attention from areas where they were concealing something to areas they wanted us to investigate, which would prove of no importance whatever. This is something that the interpreter learns only through experience. In the final analysis, the PI has to back up any report he puts out with something specific, or he has to qualify his statements. If he can't be certain what it is, then he must say so; (he can't say "I think," or "It looks like"). If he doesn't know, then the area will probably be followed up, for further investigation, in terms of priority. They may fly lower reconnaissance or use larger scale photography to see if we can extract necessary information proof to confirm our suspicions. We are getting involved in the problem of underground installations. This is a rough one. This is nothing new; it developed in World War II. The Germans had built a whole aircraft factory under a mountain, and were using the top of the mountain for a runway. It was in a crazy place and there were very few indications of what was happening. The only clues were that they had built some roads and railroad spurs into an area where there was no reason for their existence. By very careful examination of aerial photography, you started to see ventilators which stood out from trees.(fortunately in the winter). If there had been leaves there would have been no clue at all

This is true of missile sites. As you know, we are going underground and there will be very little for the interpreter to see once the side has been completed and they start building vegetation up around it. You could put a farm house over it and move it away from over the silo when you wanted to operate it. This is very difficult to interpret. Again you have to understand how these things are built and where the logical location for them would be. There is no reason for screening the area right behind the front for intermediate range ballistic missiles. They will be somewhere about 500 miles behind the lines. You have to have some understanding of how the enemy operates, what his concepts of logistics are, etc. This is where interpretation comes in. You have to tie in knowledge with what you are seeing. It's not a matter of just flashing something at you and you've got the answer.

Capt. Burhans (TAC):

I realize that all these investigations may lead into other areas and I'm appreciative of any effort to improve PI's lot. In many cases, we are still in the Stone Age in terms of technique. These projects are highly desirable, and suggest that there are possible solutions to related areas where information can be extracted much more quickly by automation. Although automation will not answer all the problems, any automatic system which will take the leg work out of the interpretation of a lot of obvious types of targets would be highly useful. I would like an automatic film viewer which has a drawing device. I'm still cranking film by hand.

CONCLUDING REMARKS

Dr. MacLeod (RADC)

Now, I'd like to give you a closing thought. I think it's obvious truism that the human being is going to continue to perform a vital function in intelligence extraction from photography for a long time to come. It follows, of course, that anything that we can do in developing our research will aid him in the key function that he will perform and is going to be very valuable. I think we've made a beginning effort; the papers presented today by the Air Force, Army and industry certainly are slanted in the right direction.

Now there are several things we are going to require to improve this effort, to give it direction and guidance so that it will really do us some good. First of all, we need ideas, we need people with sophistication and understanding, who can help us formulate problems. Important problems that require solutions to expedite either general or specific situations applying to human interpretation. We also need resources. We need, for example, to establish a repository of test photography so any researcher can get what he needs from this. I think that we've found that the difficulties in acquiring this kind of material have proved one of the biggest stumbling blocks in our efforts.

Dr. Schwartz (APRO)

In association with the interest in the repository, does your organization at this time have the capability of manning such an operation?

Dr. MacLeod

No, I am expressing this as the obvious need which I think we all recognize and certainly I think that we have to come to some decision about how to establish such a repository. Finally, I think perhaps the most important thing we need is cooperation among different specialists. We don't want to turn out to be like the five blind men of Hindustan, each describing the interpreter as something different. We have among us today psychologists, optical scientists, photographic scientists; we have experts in photo interpreter training; we have photo interpreter managers; and what we really need to do is have more meetings of this type, to get such individuals together to pool their ideas so that we can, out of this, derive approaches to the photo interpretation problem which are both realistic and at the same time objective. Approaches that will actually provide us with the techniques which can be validated. We need all kinds of specialties to do this. This is a thought to leave with you. Let's hope that within the next five years the photo interpreter is going to feel the benefits of such an effort.

ATTENDANCE LIST

SYMPOSIUM ON
HUMAN FACTOR ASPECTS OF PHOTO INTERPRETATION
HELD AT RADC, GAFB, NY
7-8 November 1962

Name	*Organization*
Mr. Attaya	ITEK Corporation
Mr. Barry	Marquardt Corporation
Maj. Becker	US Army Intelligence Center
Dr. Beechler	International Business Machines Corp.
Capt. Bendel	Air Training Command
Mr. Brock	ITEK Corporation
Mr. Burgess	Autometric Corporation
Capt. Burhans	Tactical Air Command
Dr. Chalmers	Applied Psychology Corporation
Mr. Cook	Minneapolis-Honeywell Regulator Co.
Mr. Devoe	University of Michigan
Mr. Dunker	Aeronautical Systems Division
Mr. Farina	Radio Corporation of America
Mr. Friberg	Department of Defense
Dr. Fulcher	Department of Defense
Mr. Graves	Radio Corporation of America
Mr. Gribben	Autometric Corporation
Mr. Hansen	Nortronics
Mr. Harris	Boeing Company
Mr. Hauser	Department of Army
Mr. Hoagbin	University of Michigan
Mr. Kause	Goodyear Aircraft
Dr. Klingberg	Boeing Company
Dr. Kraft	Boeing Company
Dr. Krumm	Radio Corporation of America
Mr. Lerner	Applied Psychology Corporation
Mr. Meeker	Minneapolis-Honeywell Regulator Co.
Capt. Millspaugh	Air Training Command
Mr. Pickering	Eastman Kodak Company
Dr. Porter	Applied Psychology Corporation
Mr. Rabben	Institute for Defense Analyses

Name	*Organization*
Dr. Roetling	Cornell Aeronautical Laboratory
Dr. Sadacca	US Army Personnel Research Office
Dr. Schwartz	US Army Personnel Research Office
Dr. Sinaiko	Institute for Defense Analyses
Mr. Sorem	Eastman Kodak Company
Mr. Speer	Houston-Fearless Corporation
Mr. Taub	Cornell Aeronautical Laboratory
Mr. Thomas	Boeing Company
Dr. Winterstein	International Business Machines Corp.
Mr. Ziegler	Cornell Aeronautical Laboratory

RADC PERSONNEL

Name	*Symbol*
Capt. Cavanaugh	RASHI
Lt. Col. Conklin	RASH
Dr. Crocetti	RASH
Mr. Frate	RAWIP
Capt.. Gamble	RAWIC
Capt. Hovey	RAWIO
Mr. Hughes	RAWIO
Mr. Leavitt	RASHI
Mr. Lemoine	RAWIO
Dr. MacLeod (Program Director)	RASHI
Mr. Maier	RASSL
Mr. Marks	RAWIO
Mr. Miullo	RAWIO
Mr. Pohorenec	RAWIC
Mr. Shearer	RAWIC
Mr. Sinnamon	RAWIC
2/Lt. Stenstrom	RAWIO
Mr. Stromick	RAWIP
Mr. Zieno	RAWIC